# Applied Probability

Nathanaël Berestycki and Perla Sousi*

March 18, 2015

## Contents

**1 Basic aspects of continuous time Markov chains** **3**

1.1 Markov property . . . 3

1.2 Regular jump chain . . . 4

1.3 Holding times . . . 6

1.4 Poisson process . . . 6

1.5 Birth process . . . 10

1.6 Construction of continuous time Markov chains . . . 12

1.7 Kolmogorov's forward and backward equations . . . 15

1.7.1 Countable state space . . . 15

1.7.2 Finite state space . . . 17

1.8 Non-minimal chains . . . 19

**2 Qualitative properties of continuous time Markov chains** **20**

2.1 Class structure . . . 20

2.2 Hitting times . . . 21

2.3 Recurrence and transience . . . 22

2.4 Invariant distributions . . . 23

2.5 Convergence to equilibrium . . . 27

2.6 Reversibility . . . 28

2.7 Ergodic theorem . . . 31

*University of Cambridge

**3 Queueing theory** 33

3.1 Introduction . . . 33

3.2 $M/M/1$ queue . . . 34

3.3 $M/M/\infty$ queue . . . 35

3.4 Burke's theorem . . . 36

3.5 Queues in tandem . . . 36

3.6 Jackson Networks . . . 38

3.7 Non-Markov queues: the $M/G/1$ queue. . . . 41

**4 Renewal Theory** 44

4.1 Introduction . . . 44

4.2 Elementary renewal theorem . . . 45

4.3 Size biased picking . . . 45

4.4 Equilibrium theory of renewal processes . . . 46

4.5 Renewal-Reward processes . . . 49

4.6 Example: Alternating Renewal process . . . 50

4.7 Example: busy periods of $M/G/1$ queue . . . 50

4.8 Little's formula . . . 51

4.9 $G/G/1$ queue . . . 52

**5 Population genetics** 53

5.1 Introduction . . . 53

5.2 Moran model . . . 54

5.3 Fixation . . . 54

5.4 The infinite sites model of mutations . . . 55

5.5 Kingman's $n$-coalescent . . . 56

5.6 Consistency and Kingman's infinite coalescent . . . 58

5.7 Intermezzo: Pólya's urn and Hoppe's urn* . . . 59

5.8 Infinite Alleles Model . . . 60

5.9 Ewens sampling formula . . . 61

5.10 The Chinese restaurant process . . . 64

# 1 Basic aspects of continuous time Markov chains

## 1.1 Markov property

(Most parts here are based on [1] and [2].)

A sequence of random variables is called a stochastic process or simply process. We will always deal with a countable state space $S$ and all our processes will take values in $S$.

A process is said to have the *Markov property* when the future and the past are independent given the present.

We recall now the definition of a discrete-time Markov chain.

**Definition 1.1.** The process $X = (X_n)_{n\in\mathbb{N}}$ is called a *discrete-time Markov chain* with state space $S$ if for all $x_0, \dots, x_n \in S$ we have

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

whenever both sides are well defined.

If $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ is independent of $n$, then the chain is called *time homogeneous*. We then write $P = (p_{xy})_{x,y\in S}$ for the transition matrix, i.e.

$$p_{xy} = \mathbb{P}(X_{n+1} = y \mid X_n = x)\,.$$

It is called a stochastic matrix, because it satisfies $\sum_y p_{xy} = 1$ and $p_{xy} \geq 0$ for all $x, y$.

The basic data associated to every Markov chain is the transition matrix and the starting distribution, $\mu_0$, i.e.

$$\mathbb{P}(X_0 = x_0) = \mu(x_0) \quad \text{for all} \quad x_0 \in S.$$

**Definition 1.2.** The process $X = (X_t)_{t\geq 0}$ is called a *continuous-time Markov chain* if for all $x_1, \dots, x_n \in S$ and all times $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ we have

$$\mathbb{P}\big(X_{t_n} = x_n \bigm| X_{t_{n-1}} = x_{n-1}, \dots, X_{t_1} = x_1\big) = \mathbb{P}\big(X_{t_n} = x_n \bigm| X_{t_{n-1}} = x_{n-1}\big)$$

whenever both sides are well defined.

If the right hand-side above only depends on the difference $t_n - t_{n-1}$, then the chain is called *time homogeneous.*

We write $P(t) = (p_{xy}(t))_{x,y\in S}$, where

$$p_{xy}(t) = \mathbb{P}(X_t = y \mid X_0 = x)\,.$$

The family $(P(t))_{t\geq 0}$ is called the *transition semigroup* of the continuous-time Markov chain. It is the continuous time analogue of the iterates of the transition matrix in discrete time. In the same way as in discrete time we can prove the Chapman-Kolmogorov equations for all $x, y$

$$p_{xy}(t + s) = \sum_z p_{xz}(t)p_{zy}(s).$$

Hence the transition semigroup associated to a continuous time Markov chain satisfies

- $P(t)$ is a stochastic matrix for all $t$.
- $P(t + s) = P(t)P(s)$ for all $s, t$.
- $P(0) = I$.

### 1.2 Regular jump chain

When we consider continuous time processes, there are subtleties that do not appear in discrete time. For instance if we have a disjoint countable collection of sets $(A_n)$, then

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

However, for an uncountable union $\cup_{t\geq 0} A_t$ we cannot necessarily define its probability, since it might not belong to the $\sigma$-algebra. For this reason, from now on, whenever we consider continuous time processes, we will always take them to be right-continuous, which means that for all $\omega \in \Omega$, for all $t \geq 0$, there exists $\varepsilon > 0$ (depending on $t$ and $\omega$) such that

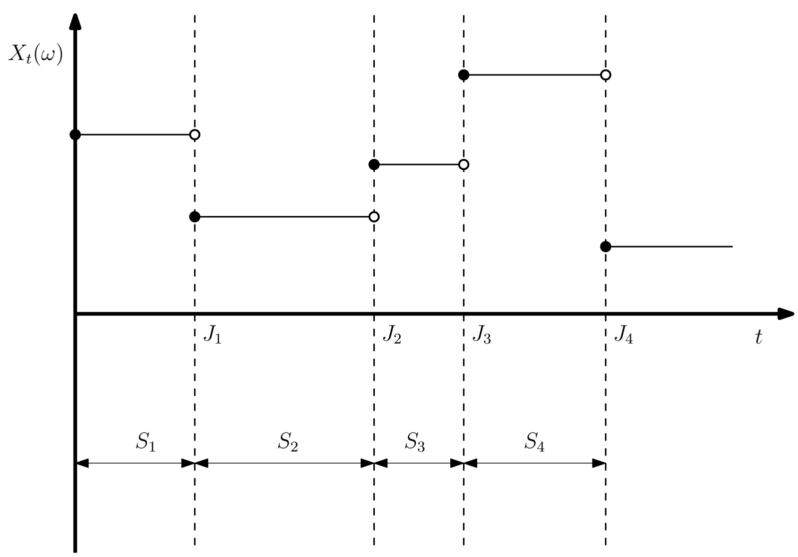
$$X_t(\omega) = X_{t+s}(\omega) \quad \text{for all } s \in [0, \varepsilon].$$

Then it follows that any probability measure concerning the process can be determined from the finite dimensional marginals $\mathbb{P}(X_{t_1} = x_1, \ldots, X_{t_n} = x_n)$.

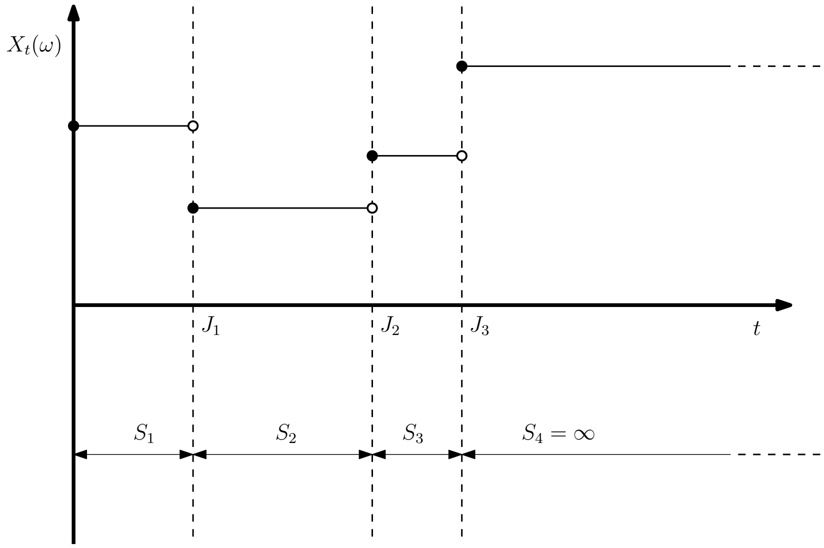
A right continuous process can have at most a countable number of discontinuities.

For every $\omega \in \Omega$ the path $t \mapsto X_t(\omega)$ of a right continuous process stays constant for a while. Three possible scenarios could arise.

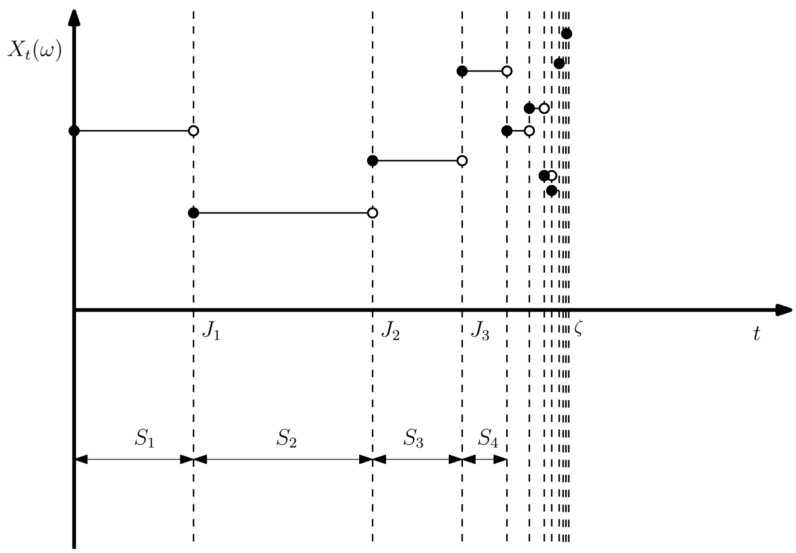
The process can make infinitely many jumps but only finitely many in every interval as shown in the figure below.




The process can get absorbed at some state as in the figure below.




The process can make infinitely many jumps in a finite interval as shown below.




We define the *jump times* $J_0, J_1, \dots$ of the continuous time Markov chain $(X_t)$ via

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\} \quad \forall\, n \geq 0,$$

where $\inf \varnothing = \infty$. We also define the *holding times* $S_1, S_2, \dots$ via

$$S_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise} \end{cases}.$$

By right continuity we get $S_n > 0$ for all $n$. If $J_{n+1} = \infty$ for some $n$, then we set $X_\infty = X_{J_n}$ as the final value, otherwise $X_\infty$ is not defined. We define the *explosion time* $\zeta$ to be

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

We also define the *jump chain* of $(X_t)$ by setting $Y_n = X_{J_n}$ for all $n$. We are not going to consider what happens to a chain after explosion. We thus set $X_t = \infty$ for $t \geq \zeta$. We call such a chain *minimal*.

## 1.3 Holding times

Let $X$ be a continuous time Markov chain on a countable state space $S$. The first question we ask is how long it stays at a state $x$. We call $S_x$ the *holding time* at $x$. Then since we take $X$ to be right-continuous, it follows that $S_x > 0$. Let $s, t \geq 0$. We now have

$$\begin{aligned}\mathbb{P}(S_x > t+s \mid S_x > s) &= \mathbb{P}(X_u = x, \ \ \forall u \in [0, t+s] \mid X_u = x, \ \ \forall u \in [0, s]) \\ &= \mathbb{P}(X_u = x, \ \ \forall u \in [s, t+s] \mid X_u = x, \ \ \forall u \in [0, s]) \\ &= \mathbb{P}(X_u = x, \ \ \forall u \in [0, t] \mid X_0 = x) = \mathbb{P}(S_x > t) \, .\end{aligned}$$

Note that the third equality follows from time homogeneity. We thus see that $S_x$ has the memoryless property. From the following theorem we get that $S_x$ has the exponential distribution and we call its parameter $q_x$.

**Theorem 1.3** (Memoryless property). *Let $S$ be a positive random variable. Then $S$ has the memoryless property, i.e.*

$$\mathbb{P}(S > t+s \mid S > s) = \mathbb{P}(S > t) \quad \forall \, s, t \geq 0$$

*if and only if $S$ has the exponential distribution.*

**Proof.** It is obvious that if $S$ has the exponential distribution, then it satisfies the memoryless property. We prove the converse. Set $F(t) = \mathbb{P}(S > t)$. Then by the assumption we get

$$F(t+s) = F(t)F(s) \quad \forall s, t \geq 0.$$

Since $S > 0$, there exists $n$ large enough so that $F(1/n) = \mathbb{P}(S > 1/n) > 0$. We thus obtain

$$F(1) = F\left(\frac{1}{n} + \ldots + \frac{1}{n}\right) = F\left(\frac{1}{n}\right)^n > 0,$$

and hence we can set $F(1) = e^{-\lambda}$ for some $\lambda \geq 0$. It now follows that for all $k \in \mathbb{N}$ we have

$$F(k) = e^{-\lambda k}.$$

Similarly for all rational numbers we get

$$F(p/q) = F(1/q)^p = F(1)^{p/q} = e^{-\lambda p/q}.$$

It remains to show that the same is true for all $t \in \mathbb{R}_+$. But for each such $t$ and $\varepsilon > 0$ we can find rational numbers $r, s$ so that $r \leq t \leq s$ and $|r - s| \leq \varepsilon$. Since $F$ is decreasing, we deduce

$$e^{-\lambda s} \leq F(t) \leq e^{-\lambda r}.$$

Taking $\varepsilon \to 0$ finishes the proof. □

## 1.4 Poisson process

We are now going to look at the simplest example of a continuous-time Markov chain, the Poisson process.

Suppose that $S_1, S_2, \ldots$ are i.i.d. random variables with $S_1 \sim \text{Exp}(\lambda)$. Define the jump times $J_1 = S_1$ and for all $n$ define $J_n = S_1 + \ldots + S_n$ and set $X_t = i$ if $J_i \leq t < J_{i+1}$. Then $X$ is called a *Poisson process* of parameter $\lambda$.

**Theorem 1.4** (Markov property)**.** *Let $(X_t)_{t\geq 0}$ be a Poisson process of rate $\lambda$. Then for all $s \geq 0$ the process $(X_{s+t} - X_s)_{t\geq 0}$ is also a Poisson process of rate $\lambda$ and is independent of $(X_r)_{r\leq s}$.*

**Proof.** We set $Y_t = X_{t+s} - X_s$ for all $t \geq 0$. Then

$$Y_t = \sum_{i=1}^{\infty} \mathbf{1}(s < J_i \leq t + s). \tag{1.1}$$

From this it follows that it suffices to show that $Y_t$ is independent of the event $\{X_s = k\}$ for all $k$. From the definition of $X$ we have $\{X_s = k\} = \{J_k \leq s\} \cap \{S_{k+1} > s - J_k\}$. Therefore in the sum in (1.1) we only keep the terms from $k+1$ onwards. Since $J_i = J_{i-1} + S_i$ for all $i$, by conditioning on $S_{k+1} > s - J_k$, using the memoryless property of the exponential distribution, we see that the holding times for $Y$ are again Exponential random variables with parameter $\lambda$ and independent of $\{X_s\}$. □

Similarly to the proof above, one can show the strong Markov property for the Poisson process.

Recall from the discrete setting that a random variable $T$ with values in $[0, \infty]$ is called a *stopping time* if the event $\{T \leq t\}$ depends on $(X_s)_{s\leq t}$ for all $t$.

**Theorem 1.5** (Strong Markov property)**.** *Let $(X_t)_{t\geq 0}$ be a Poisson process with rate $\lambda$ and let $T$ be a stopping time. Then conditional on $T < \infty$, the process $(X_{T+t} - X_T)_{t\geq 0}$ is also a Poisson process of rate $\lambda$ and is independent of $(X_s)_{s\leq T}$.*

The following theorem gives two equivalent characterizations of the Poisson process.

**Theorem 1.6.** *Let $(X_t)$ be an increasing right-continuous process taking values in $\{0, 1, 2, \ldots\}$ with $X_0 = 0$. Let $\lambda > 0$. Then the following statements are equivalent:*

(a) *The holding times $S_1, S_2, \ldots$ are i.i.d. Exponentially distributed with parameter $\lambda$ and the jump chain is given by $Y_n = n$, i.e. $X$ is a Poisson process.*

(b) [infinitesimal] *$X$ has independent increments and as $h \downarrow 0$ uniformly in $t$ we have*

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

(c) *$X$ has independent and stationary increments and for all $t \geq 0$ we have $X_t \sim$Poisson$(\lambda t)$.*

**Proof.** (a)⇒(b) If (a) holds, then by the Markov property the increments are independent and stationary. Using stationarity we have uniformly in $t$ as $h \to 0$

$$\mathbb{P}(X_{t+h} - X_t = 0) = \mathbb{P}(X_h = 0) = \mathbb{P}(S_1 > h) = e^{-\lambda h} = 1 - \lambda h + o(h)$$
$$\mathbb{P}(X_{t+h} - X_t \geq 1) = \mathbb{P}(X_h \geq 1) = \mathbb{P}(S_1 \leq h) = 1 - e^{-\lambda h} = \lambda h + o(h)$$
$$\mathbb{P}(X_{t+h} - X_t \geq 2) \leq \mathbb{P}(S_1 \leq h, S_2 \leq h) = (1 - e^{-\lambda h})^2 = o(h),$$

which means that (b) holds.

(b)⇒(c) We set $p_j(t) = \mathbb{P}(X_t = j)$. Since the increments are independent and $X$ is increasing, we get

$$p_j(t+h) = \sum_{i=0}^{j} \mathbb{P}(X_t = j - i)\,\mathbb{P}(X_{t+h} - X_t = i) = p_j(t)(1 - \lambda h + o(h)) + p_{j-1}(t)(\lambda h + o(h)) + o(h).$$

From this it follows that $p_j(t)$ is continuous and rearranging we obtain

$$\frac{p_j(t+h) - p_j(t)}{h} = -\lambda p_j(t) + \lambda p_{j-1}(t) + O(h). \tag{1.2}$$

Since this holds uniformly in $t$ we can set $s = t + h$ and get

$$\frac{p_j(s) - p_j(s-h)}{h} = -\lambda p_j(s-h) + \lambda p_{j-1}(s-h) + O(h). \tag{1.3}$$

Therefore, from (1.2) and (1.3) we see that we can take the limit as $h \to 0$ and get

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t).$$

By differentiating $e^{\lambda t} p_k(t)$ with respect to $t$ and substituting the above we get

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

For $j = 0$, since in order to be at 0 at time $t + h$ we must be at 0 at time $t$, it follows similarly to above that

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

which gives that $p_0(t) = e^{-\lambda t}$. We thus get $p_1(t) = e^{-\lambda t} \lambda t$ and inductively we obtain

$$p_n(t) = e^{-\lambda t} (\lambda t)^n / n!.$$

It follows that $X_t$ has the Poisson distribution with parameter $\lambda t$. If $X$ satisfies (b), then $(X_{t+s} - X_s)_t$ also satisfies (b), and hence $X$ has stationary increments.

(c)⇒(a) The Poisson process satisfies (c). But (c) determines uniquely the finite dimensional marginals for a right-continuous and increasing process. Therefore (c) implies (a). □

**Theorem 1.7** (Superposition)**.** *Let $X$ and $Y$ be two independent Poisson processes with parameters $\lambda$ and $\mu$ respectively. Then $Z_t = X_t + Y_t$ is also a Poisson process with parameter $\lambda + \mu$.*

**Proof.** We are going to use the infinitesimal definition of a Poisson process. Using the independence of $X$ and $Y$ we get uniformly in $t$ as $h \to 0$

$$\begin{aligned}\mathbb{P}(Z_{t+h} - Z_t = 0) &= \mathbb{P}(X_{t+h} - X_t = 0, Y_{t+h} - Y_t = 0) = \mathbb{P}(X_{t+h} - X_t = 0)\,\mathbb{P}(Y_{t+h} - Y_t = 0) \\ &= (1 - \lambda h + o(h))(1 - \mu h + o(h)) = 1 - (\lambda + \mu)h + o(h).\end{aligned}$$

We also have uniformly in $t$ as $h \to 0$

$$\begin{aligned}\mathbb{P}(Z_{t+h} - Z_t = 1) &= \mathbb{P}(X_{t+h} - X_t = 0, Y_{t+h} - Y_t = 1) + \mathbb{P}(X_{t+h} - X_t = 1, Y_{t+h} - Y_t = 0) \\ &= (1 - \lambda h + o(h))(\mu h + o(h)) + (1 - \mu h + o(h))(\lambda h + o(h)) = (\lambda + \mu)h + o(h).\end{aligned}$$

Clearly $Z$ has independent increments, and hence it is a Poisson process of parameter $\lambda + \mu$. □

**Theorem 1.8** (Thinning)**.** *Let $X$ be a Poisson process of parameter $\lambda$. Let $(Z_i)_i$ be i.i.d. Bernoulli random variables with success probability $p$. Let $Y$ be a process with values in $\mathbb{N}$, which jumps at time $t$ if and only if $X$ jumps and $Z_{X_t} = 1$. In other words, we keep every point of $X$ with probability $p$ independently over different points. Then $Y$ is a Poisson process of parameter $\lambda p$ and $X - Y$ is an independent Poisson process of parameter $\lambda(1-p)$.*

**Proof.** We will use the infinitesimal definition of a Poisson process to prove the result. The independence of increments for $Y$ is clear. Since $\mathbb{P}(X_{t+h} - X_t \geq 2) = o(h)$ uniformly for all $t$ as $h \to 0$ we have

$$\begin{aligned}\mathbb{P}(Y_{t+h} - Y_t = 1) &= p\mathbb{P}(X_{t+h} - X_t = 1) + o(h) = \lambda ph + o(h)\\ \mathbb{P}(Y_{t+h} - Y_t = 0) &= \mathbb{P}(X_{t+h} - X_t = 0) + \mathbb{P}(X_{t+h} - X_t = 1)\,(1-p) + o(h)\\ &= 1 - \lambda h + o(h) + (\lambda h + o(h))(1-p) = 1 - \lambda ph + o(h).\end{aligned}$$

This proves that $Y$ is a Poisson process of rate $\lambda p$. Since $X - Y$ is a thinning of $X$ with probability $1-p$, it follows from above that it is also a Poisson process of rate $\lambda(1-p)$. To prove the independence, since both processes are right-continuous and increasing, it is enough to check the finite dimensional marginals are independent, i.e. that for $t_1 \leq t_2 \leq \ldots \leq t_k$ and $n_1 \leq \ldots \leq n_k$, $m_1 \leq \ldots \leq m_k$

$$\begin{aligned}&\mathbb{P}(Y_{t_1} = n_1, \ldots, Y_{t_k} = n_k, X_{t_1} - Y_{t_1} = m_1, \ldots, X_{t_k} - Y_{t_k} = m_k)\\ &= \mathbb{P}(Y_{t_1} = n_1, \ldots, Y_{t_k} = n_k)\,\mathbb{P}(X_{t_1} - Y_{t_1} = m_1, \ldots, X_{t_k} - Y_{t_k} = m_k)\,.\end{aligned}$$

We will only show it for a fixed $t$, but the general case follows similarly. Using the independence of the $Z_i$'s we now get

$$\begin{aligned}\mathbb{P}(Y_t = n, X_t - Y_t = m) = \mathbb{P}(X_t = m+n, Y_t = n) &= e^{-\lambda}\frac{\lambda^{\lambda+\mu}}{(m+n)!}\binom{m+n}{n}p^n(1-p)^m\\ &= \left(e^{-\lambda p}\frac{(\lambda p)^n}{n!}\right)\left(e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}\right) = \mathbb{P}(Y_t = n)\,\mathbb{P}(X_t - Y_t = m)\end{aligned}$$

which shows that $Y_t$ is independent of $X_t - Y_t$ for all times $t$. □

**Theorem 1.9.** *Let $X$ be a Poisson process. Conditional on the event $\{X_t = n\}$ the jump times $J_1, \ldots, J_n$ have joint density function*

$$f(t_1, \ldots, t_n) = \frac{n!}{t^n}\mathbf{1}(0 \leq t_1 \leq \ldots \leq t_n \leq t).$$

**Remark 1.10.** First we notice that the joint density given in the theorem above is the same as the one for an ordered sample from the uniform distribution on $[0, t]$. Also, let's check that it is indeed a density. Take for instance $n = 2$. Then the integral over the region $0 \leq t_1 \leq t_2 \leq t$ is the area of the triangle, which is given by $t^2/2$. More generally, for all $n$, if we integrate the function $1/t^n$ over the cube $[0, t]^n$, then it is equal to 1. By symmetry, when integrating over $t_1 \leq t_2 \leq \ldots \leq t_n$ we have to divide the whole area by $n!$.

**Proof of Theorem 1.9.** Since the times $S_1.S_2, \ldots, S_{n+1}$ are independent, their joint density is given by

$$\lambda^{n+1}e^{-\lambda(s_1+\ldots+s_{n+1})}\mathbf{1}(s_1, \ldots, s_{n+1} \geq 0).$$

The jump times $J_1 = S_1, J_2 = S_1 + S_2, \ldots, J_{n+1} = S_1 + \ldots + S_{n+1}$ have joint density given by

$$\lambda^{n+1}e^{-\lambda t_{n+1}}\mathbf{1}(0 \leq t_1 \leq \ldots \leq t_{n+1}). \tag{1.4}$$

This is easy to check by induction on $n$. For $n = 2$ we have

$$\mathbb{P}(S_1 \leq x_1, S_1 + S_2 \leq x_2) = \int_0^{x_1} \lambda e^{-\lambda x}\mathbb{P}(S_2 \leq x_2 - x)\;dx,$$

and hence by differentiating we obtain the formula above. The rest follows using the same idea. Another way to show (1.4) is by using the formula for the density of a transformation of a random vector (see notes of Probability 1A).

Now take $A \subseteq \mathbb{R}^n$. Then

$$\mathbb{P}((J_1, \ldots, J_n) \in A \mid X_t = n) = \frac{\mathbb{P}((J_1, \ldots, J_n) \in A, X_t = n)}{\mathbb{P}(X_t = n)}.$$

For the numerator using the density above we obtain

$$\begin{aligned}
&\mathbb{P}((J_1, \ldots, J_n) \in A, X_t = n) = \mathbb{P}((J_1, \ldots, J_n) \in A, J_n \le t < J_{n+1}) \\
&= \int_{(t_1,\ldots,t_n)\in A} \int_t^\infty \lambda^{n+1} e^{-\lambda t_{n+1}} \mathbf{1}(0 \le t_1 \le \ldots \le t_n \le t \le t_{n+1})\, dt_1 \ldots dt_n\, dt_{n+1} \\
&= \int_{(t_1,\ldots,t_n)\in A} \lambda^n e^{-\lambda t} \mathbf{1}(0 \le t_1 \le \ldots \le t_n \le t)\, dt_1 \ldots dt_n \\
&= \lambda^n e^{-\lambda t} \int_{(t_1,\ldots,t_n)\in A} \mathbf{1}(0 \le t_1 \le \ldots \le t_n \le t)\, dt_1 \ldots dt_n.
\end{aligned}$$

Since $X_t$ has the Poisson distribution of parameter $\lambda t$, we get $\mathbb{P}(X_t = n) = (\lambda t)^n e^{-\lambda t}/n!$. Dividing the two expressions we obtain

$$\mathbb{P}((J_1, \ldots, J_n) \in A \mid X_t = n) = \frac{n!}{t^n} \int_{(t_1,\ldots,t_n)\in A} \mathbf{1}(0 \le t_1 \le \ldots \le t_n \le t)\, dt_1 \ldots dt_n,$$

and hence the joint density of the jump times $J_1, \ldots, J_n$ given $X_t = n$ is equal to $f$. □

## 1.5 Birth process

The birth process is a generalization of the Poisson process in which the parameter $\lambda$ can depend on the current state of the process. For the Poisson process the rate of going from $i$ to $i+1$ is $\lambda$. For the birth process this rate is equal to $q_i$. For each $i$ let $S_i$ be an Exponential random variable with parameter $q_i$. Suppose that $S_1, S_2, \ldots$ are independent. Set $J_i = S_1 + \ldots + S_i$ and $X_t = i$ whenever $J_i \le t < J_{i+1}$. Then $X$ is called a *birth process*.

### Simple birth process

Suppose now that for all $i$ we have $q_i = \lambda i$. We can think of this particular birth process as follows: at time 0 there is only one individual, i.e. $X_0 = 1$. Each individual has an exponential clock of parameter $\lambda$. We will now need a standard result about Exponential random variables, whose proof you are asked to give in the example sheet.

**Proposition 1.11.** *Let $(T_k)_{k\ge1}$ be a sequence of independent random variables with $T_k \sim \mathrm{Exp(q_k)}$ and $0 < q = \sum_k q_k < \infty$. Set $T = \inf_k T_k$. Then this infimum is attained at a unique point $K$ with probability 1. Moreover, the random variables $T$ and $K$ are independent, with $T \sim \mathrm{Exp(q)}$ and $\mathbb{P}(K = k) = q_k/q$.*

Using the result above it follows that if there are $i$ individuals, then the first clock will ring after an exponential time of parameter $\lambda i$. Then we have $i + 1$ individuals and by the memoryless property of the exponential distribution, the process begins afresh. Let $X_t$ denote the number of individuals at time $t$ when $X_0 = 1$. We want to find $\mathbb{E}[X_t]$.

A probabilistic way to calculate it is the following: let $T$ be the time that the first birth takes place. Then we have

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[X_t\mathbf{1}(T \le t)] + \mathbb{E}[X_t\mathbf{1}(T > t)] = \mathbb{E}[X_t\mathbf{1}(T \le t)] + e^{-\lambda t}\\ &= \int_0^t \lambda e^{-\lambda s}\mathbb{E}[X_t \mid T = s]\, ds + e^{-\lambda t}.\end{aligned}$$

If we set $\mu(t) = \mathbb{E}[X_t]$, then by the memoryless property of the exponential distribution and the fact that at time $T$ there are 2 individuals, we get that $\mathbb{E}[X_t \mid T = s] = 2\mu(t-s)$. So we deduce

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

and by a change of variable in the integral we get

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

By differentiating we obtain

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

and hence $\mu(t) = e^{\lambda t}$.

Most of the theory of Poisson processes goes through to birth processes without much difference. The main difference between the two processes is the possibility of explosion in birth processes. For a general birth process with rates $(q_i)$ we have that $S_i$ is an exponential random variable of parameter $q_i$ and the jump times are $J_1 = S_1$ and $J_n = S_1 + \ldots + S_n$ for all $n$. Explosion occurs when $\zeta = \sup_n J_n = \sum_n S_n < \infty$.

**Proposition 1.12.** *Let $X$ be a birth process with rates $(q_i)$ with $X_0 = 1$.*

(1) *If $\sum_{i=1}^{\infty} 1/q_i < \infty$, then $\mathbb{P}(\zeta < \infty) = 1$.*

(2) *If $\sum_{i=1}^{\infty} 1/q_i = \infty$, then $\mathbb{P}(\zeta = \infty) = 1$.*

**Proof.** (1) If $\sum_n 1/q_n < \infty$, then we get $\mathbb{E}[\sum_n S_n] = \sum_n 1/q_n < \infty$, and hence $\mathbb{P}(\sum_n S_n < \infty) = 1$.

(2) If $\sum_n 1/q_n = \infty$, then $\prod_n (1 + 1/q_n) = \infty$. By monotone convergence and independence we get

$$\mathbb{E}\left[\exp\left(-\sum_{n=1}^{\infty} S_n\right)\right] = \prod_{n=1}^{\infty}\mathbb{E}[\exp(-S_n)] = \prod_{n=1}^{\infty}\left(1 + \frac{1}{q_n}\right)^{-1} = 0.$$

Hence this gives that $\sum_n S_n = \infty$ with probability 1. □

**Theorem 1.13** (Markov property)**.** *Let $X$ be a birth process of rates $(q_i)$. Then conditional on $X_s = i$, the process $(X_{s+t})_{t\ge 0}$ is a birth process of rates $(q_i)$ starting from $i$ and independent of $(X_r : r \le s)$.*

The proof of the Markov property follows similarly to the case of the Poisson process.

**Theorem 1.14.** *Let $X$ be an increasing, right-continuous process with values in $\{1, 2, \ldots\} \cup \{\infty\}$. Let $0 \le q_j < \infty$ for all $j \ge 0$. Then the following three conditions are equivalent:*

(1) *(jump chain/holding time definition) condition on $X_0 = i$, the holding times $S_1, S_2, \dots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \dots$ respectively and the jump chain is given by $Y_n = i + n$ for all $n$;*

(2) *(infinitesimal definition) for all $t, h \geq 0$, conditional on $X_t = i$ the process $(X_{t+h})_j$ is independent of $(X_s : s \leq t)$ and as $h \downarrow 0$ uniformly in $t$*

$$\mathbb{P}(X_{t+h} = i \mid X_t = i) = 1 - q_i h + o(h),$$
$$\mathbb{P}(X_{t+h} = i + 1 \mid X_t = i) = q_i h + o(h);$$

(3) *(transition probability definition) for all $n = 0, 1, 2, \dots$, all times $0 \leq t_0 \leq \dots \leq t_{n+1}$ and all states $i_0, \dots, i_{n+1}$*

$$\mathbb{P}\big(X_{t_{n+1}} = i_{n+1} \bigm| X_0 = i_0, \dots, X_{t_n} = i_n\big) = p_{i_n i_{n+1}}(t_{n+1} - t_n),$$

*where $(p_{i,j}(t) : i, j = 0, 1, 2, \dots)$ is the unique solution of the forward equations $P'(t) = P(t)Q$.*

### 1.6 Construction of continuous time Markov chains

In this section we are going to give three probabilistic constructions of a continuous time Markov chain. We start with the definition of a $Q$-matrix.

**Definition 1.15.** Let $S$ be a countable set. Then a $Q$-*matrix* on $S$ is a matrix $Q = (q_{ij} : i, j \in S)$ satisfying the following:

- $0 \leq -q_{ii} < \infty$ for all $i$;
- $q_{ij} \geq 0$ for all $i \neq j$;
- $\sum_j q_{ij} = 0$ for all $i$.

We define $q_i := -q_{ii}$ for all $i \in S$. Given a $Q$-matrix $Q$ we define a jump matrix as follows: for $x \neq y$ with $q_x \neq 0$ we set

$$p_{xy} = \frac{q_{xy}}{-q_{xx}} = \frac{q_{xy}}{q_x} \quad \text{and} \quad p_{xx} = 0.$$

If $q_x = 0$, then we set $p_{xy} = \mathbf{1}(x = y)$.

From this definition, since $Q$ is a $Q$-matrix, it immediately follows that $P$ is a stochastic matrix.

**Example 1.16.** Suppose that $S = \{1, 2, 3\}$ and let

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

Then the associated jump matrix $P$ is given by

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Recall the definition of a minimal chain to be one that is set equal to $\infty$ after explosion. From now on we consider minimal chains.

**Definition 1.17.** A Markov chain $X$ with initial distribution $\lambda$ and infinitesimal generator $Q$, which is a $Q$-matrix, is a stochastic process with jump chain $Y_n = X_{J_n}$ being a discrete time Markov chain with $Y_0 \sim \lambda$ and transition matrix $P$ such that conditional on $Y_0, Y_1, \ldots, Y_n$ the holding times $S_1, S_2, \ldots, S_{n+1}$ are independent Exponential random variables with rates $q(Y_0), \ldots, q(Y_n)$. We write $(X_t) = \text{Markov}(Q, \lambda)$.

We now give three constructions of a Markov chain with generator $Q$.

**Construction 1**

- Take $Y$ a discrete time Markov chain with initial distribution $\lambda$ and transition matrix $P$.
- Take $(T_i)_{i\geq 1}$ i.i.d. Exp(1), independent of $Y$ and set $S_n = \frac{T_n}{q(Y_{n-1})}$ and $J_n = \sum_{i=1}^n S_i$.
- Set $X_t = Y_n$ if $J_n \leq t < J_{n+1}$ and $X_t = \infty$ otherwise.

Note that this construction satisfies the definition, because if $T \sim \text{Exp}(1)$, then $T/\mu \sim \text{Exp}(\mu)$.

**Construction 2**

Let $(T_n^y)_{n\geq 1, y\in S}$ be i.i.d. Exp(1) random variables. Define inductively $Y_n, S_n$ as follows: $Y_0 \sim \lambda$ and inductively if $Y_n = x$, then we set for $y \neq x$

$$S_{n+1}^y = \frac{T_{n+1}^y}{q_{xy}} \sim \text{Exp}(q_{xy}) \quad \text{and} \quad S_{n+1} = \inf_{y\neq x} S_{n+1}^y.$$

If $q_x > 0$, then by Proposition 1.11 $S_{n+1} = S_{n+1}^Z$ for some random $Z$ in the state space and $S_{n+1} \sim \text{Exp}(q_x)$. In this case we take $Y_{n+1} = Z$. If $q_x = 0$, then we take $Y_{n+1} = x$.

**Remark 1.18.** This construction shows that the rate at which we leave a state $x$ is equal to $q_x$ and we transition to $y$ from $x$ at rate $q_{xy}$.

**Construction 3**

We consider independent Poisson processes for each pair of points $x, y$ with $x \neq y$ with parameter $q_{xy}$. We define $Y_n, J_n$ inductively as follows: first we take $Y_0 \sim \lambda$ and set $J_0 = 0$. If $Y_n = x$, then we set

$$J_{n+1} = \inf\{t > J_n : N_t^{Y_n y} \neq N_{J_n}^{Y_n y} \text{ for some } y \neq Y_n\}, \; Y_{n+1} = \begin{cases} y & \text{if } J_{n+1} < \infty \text{ and } N_{J_{n+1}}^{Y_n y} \neq N_{J_n}^{Y_n y} \\ x & \text{if } J_{n+1} = \infty \end{cases}.$$

Recall the definition of the explosion time for a jump process. If $(J_n)$ are the jump times, then the explosion time $\zeta$ is defined to be $\zeta = \sup_n J_n = \sum_{n=1}^\infty S_n$.

**Theorem 1.19.** *Let $X$ be Markov$(Q, \lambda)$ on $S$. Then $\mathbb{P}(\zeta = \infty) = 1$ if any of the following holds:*

(i) *$S$ is finite;*

(ii) *$\sup_x q_x < \infty$;*

(iii) *$X_0 = x$ and $x$ is recurrent for the jump chain.*

**Proof.** Since (i) implies (ii), we will prove $\mathbb{P}(\zeta = \infty) = 1$ under (ii). We set $q = \sup_x q_x$. The holding times satisfy $S_n \geq T_n/q$, where $T_n \sim \text{Exp}(1)$ and are i.i.d. By the strong law of large numbers we now obtain

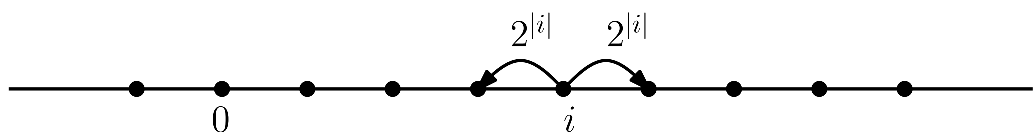
$$\zeta = \sum_{n=1}^{\infty} S_n \geq \frac{1}{q} \cdot \sum_{n=1}^{\infty} T_n = \infty \quad \text{with probability } 1.$$

Suppose that (iii) holds and let $(N_i)_{i\geq 1}$ be the times when the jump chain $Y$ visits $x$. By the strong law of large numbers again we get

$$\zeta \geq \sum_{i=1}^{\infty} S_{N_i} = \frac{1}{q_x} \cdot \sum_{i=1}^{\infty} T_{N_i} = \infty \quad \text{with probability } 1$$

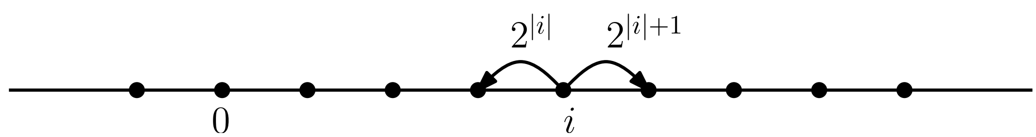
and this finishes the proof. □

**Example 1.20.** Consider a continuous time Markov chain on $\mathbb{Z}$ with rates as given in the Figure below.




Thus the jump chain is a simple symmetric random walk on $\mathbb{Z}$. Since it is recurrent, it follows that there is no explosion.

**Example 1.21.** Consider a continuous time Markov chain on $\mathbb{Z}$ with rates as given in the figure below.




Thus the jump chain is a biased random walk on $\mathbb{Z}$ with $\mathbb{P}(\xi = +1) = 2/3 = 1 - \mathbb{P}(\xi = -1)$. The expected number of times this chain visits any given vertex is at most 3, and hence

$$\mathbb{E}[\zeta] = \mathbb{E}\left[\sum_{n=1}^{\infty} S_i\right] \leq \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} < \infty,$$

which implies that $\zeta < \infty$ a.s., i.e. there is explosion.

**Theorem 1.22** (Strong Markov property)**.** *Let $X$ be Markov$(Q, \lambda)$ and let $T$ be a stopping time. Then conditional on $T < \zeta$ and $X_T = x$ the process $(X_{T+t})_{t\geq 0}$ is Markov$(Q, \delta_x)$ and independent of $(X_s)_{s\leq T}$.*

We will not give the proof here. The idea is similar to the discrete time setting, but it requires some more measure theory. For a proof we refer the reader to [2, Section 6.5].

## 1.7 Kolmogorov's forward and backward equations

### 1.7.1 Countable state space

**Theorem 1.23.** *Let $X$ be a minimal right continuous process with values in a countable set $S$. Let $Q$ be a $Q$-matrix with jump matrix $P$. Then the following conditions are equivalent:*

(a) *$X$ is a continuous time Markov chain with generator $Q$;*

(b) *for all $n \geq 0$, $0 \leq t_0 \leq \ldots \leq t_n$ and all states $x_0, \ldots, x_{n+1}$*

$$\mathbb{P}\big(X_{t_{n+1}} = x_{n+1} \bigm| X_{t_n} = x_n, \ldots, X_{t_0} = x_0\big) = p_{x_n x_{n+1}}(t_{n+1} - t_n),$$

*where $(p_{xy}(t))$ is the minimal non-negative solution to the backward equation*

$$P'(t) = QP(t) \quad \text{and} \quad P(0) = I.$$

**Remark 1.24.** Note that minimality in the above theorem refers to the fact that if $\widetilde{P}$ is another non-negative solution, then $\widetilde{p}_{xy}(t) \geq p_{xy}(t)$. It is actually related to the fact that we restrict attention to minimal chains, i.e. those that jump to a cemetery state after explosion.

**Proof of Theorem 1.23.** (a)$\Rightarrow$(b) Since $X$ has the Markov property, it immediately follows that

$$\mathbb{P}\big(X_{t_{n+1}} = x_{n+1} \bigm| X_{t_n} = x_n, \ldots, X_{t_0} = x_0\big) = \mathbb{P}(X_{tn+1} = x_{n+1} \mid X_{t_n} = x_n)\,.$$

We define $P(t)$ by setting $p_{xy}(t) = \mathbb{P}_x(X_t = y)$ for all $x, y \in S$. We show that $P(t)$ is the minimal non-negative solution to the backward equation.

If $(J_n)$ denote the jump times of the chain, we have

$$\mathbb{P}_x(X_t = y, J_1 > t) = e^{-q_x t}\mathbf{1}(x = y). \tag{1.5}$$

By integrating over the values of $J_1 \leq t$ and using the independence of the jump chain we get for $z \neq x$

$$\mathbb{P}_x(X_t = y, J_1 \leq t, X_{J_1} = z) = \int_0^t q_x e^{-q_x s} \frac{q_{xz}}{q_x} p_{t-s}(z, y)\, ds = \int_0^t e^{-q_x s} q_{xz} p_{t-s}(z, y)\, ds.$$

Taking the sum over all $z \neq x$ and using monotone convergence we obtain

$$\begin{aligned} p_{xy}(t) = \mathbb{P}_x(X_t = y) &= e^{-q_x t}\mathbf{1}(x = y) + \sum_{z \neq x} \int_0^t e^{-q_x s} q_{xz} p_{t-s}(z, y)\, ds \\ &= e^{-q_x t}\mathbf{1}(x = y) + \int_0^t \sum_{z \neq x} e^{-q_x s} q_{xz} p_{t-s}(z, y)\, ds \end{aligned}$$

or equivalently we have

$$e^{q_x t} p_{xy}(t) = \mathbf{1}(x = y) + \int_0^t \sum_{z \neq x} e^{q_x u} q_{xz} p_u(z, y)\, du.$$

From this it follows that $p_{xy}(t)$ is continuous in $t$. We now notice that $\sum_{z\neq x} q_{xz}p_u(z,y)$ is a uniformly convergent series of continuous functions, and hence the limit is continuous. Therefore, we see that the right hand side is differentiable and we deduce

$$e^{q_x t}(q_x p_{xy}(t) + p'_{xy}(t)) = \sum_{z\neq x} e^{q_x t} q_{xz} p_t(z,y).$$

Cancelling the exponential terms and rearranging we finally get

$$p'_{xy}(t) = \sum_{z} q_{xz} p_t(z,y),$$

which shows that $P'(t) = QP(t)$.

Let now $\widetilde{P}$ be another non-negative solution of the backward equations. We show that for all $x, y, t$ we have $p_{xy}(t) \leq \widetilde{p}_{xy}(t)$.

Using the same argument as before we can show that

$$\mathbb{P}_x(X_t = y, t < J_{n+1}) = e^{-q_x t}\mathbf{1}(x = y) + \sum_{z\neq x} \int_0^t e^{-q_x s} q_{xz} \mathbb{P}_z(X_{t-s} = y, t-s < J_n)\, ds. \tag{1.6}$$

If $\widetilde{P}$ satisfies the backward equations, then by reversing the steps that led to (1.7) we see that it also satisfies

$$\widetilde{p}'_{xy}(t) = e^{-q_x t}\mathbf{1}(x = y) + \sum_{z\neq x} \int_0^t e^{-q_x s} q_{xz} \widetilde{p}_{zy}(t-s)\, ds. \tag{1.7}$$

We now show by induction that $\mathbb{P}_x(X_t = y, t < J_n) \leq \widetilde{p}_{xy}(t)$ for all $n$. Indeed, for $n = 1$, it immediately follows from (1.5). Next suppose that it holds for $n$, i.e.

$$\mathbb{P}_x(X_t = y, t < J_n) \leq \widetilde{p}_{xy}(t) \quad \forall x, y, t.$$

Then using (1.6) and (1.7) we obtain that

$$\mathbb{P}_x(X_t = y, t < J_{n+1}) \leq \widetilde{p}_{xy}(t) \quad \forall x, y, t.$$

Using minimality of the chain, i.e. that after explosion it jumps to a cemetery state, we therefore conclude that for all $x, y$ and $t$

$$p_{xy}(t) = \mathbb{P}_x(X_t = y, t < \zeta) = \lim_{n\to\infty} \mathbb{P}_x(X_t = y, t < J_n) \leq \widetilde{p}_{xy}(t)$$

and this proves the minimality of $P$.

(b)⇒(a) This follows in the same way as in the Poisson process case. We already showed that if $X$ is Markov with generator $Q$, then its transition semigroup must satisfy (b). But (b) determines uniquely the finite dimensional marginals of $X$ and since $X$ is right-continuous, we get the uniqueness in law. □

**Remark 1.25.** Note that in the case of a finite state space, the above proof becomes easier, since interchanging sum and integral becomes obvious. We will see in the next section that in the finite state space case, the solution to the backward equation can be written as $P(t) = e^{tQ}$ and it is also the solution to the forward equation, i.e. $P'(t) = P(t)Q$. In the infinite state space case though, to show that the semigroup satisfies the forward equations, one needs to employ a time-reversal argument. More details on that can be found in [2, Section 2.8].

### 1.7.2 Finite state space

We now restrict attention to the case of a finite state space. In this case the solution to the backward equation has a simple expression. We start by defining the exponential of a finite dimensional matrix.

**Definition 1.26.** If $A$ is a finite-dimensional matrix, then we define its exponential via

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

**Claim 1.1.** *For any $r \times r$ matrix $A$, the exponential $e^A$ is a finite dimensional matrix too and if $A_1$ and $A_2$ commute, then*

$$e^{A_1+A_2} = e^{A_1} \cdot e^{A_2}.$$

*Sketch of Proof.* We set $\Delta = \max_{i,j} |a_{ij}|$. Then by induction it is not hard to check that for all $i, j$

$$|a_{ij}(n)| \leq r^{n-1}\Delta^n.$$

Using this, then any term of $e^A$ is bounded in absolute value by a convergent series, and hence this shows the convergence in every component of the sum appearing in the definition of $e^A$.

The commutativity property follows easily now using the definition. □

**Theorem 1.27.** *Let $Q$ be a $Q$-matrix on a finite set $S$ and let $P(t) = e^{tQ}$. Then $(P(t))_t$ has the following properties:*

(1) *$P(t + s) = P(t)P(s)$ for all $s, t$;*

(2) *$(P(t))$ is the unique solution to the forward equation $\frac{d}{dt}P(t) = P(t)Q$ and $P(0) = I$;*

(3) *$(P(t))$ is the unique solution to the backward equation $\frac{d}{dt}P(t) = QP(t)$ and $P(0) = I$;*

(4) *for $k = 0, 1, \dots$ we have $\left(\frac{d}{dt}\right)^k |_{t=0}\, P(t) = Q^k$.*

**Proof.** (1) Since the matrices $sQ$ and $tQ$ commute for all $s, t \geq 0$, the first property is immediate.

(2)-(3) Since the sum defining $e^{tQ}$ has infinite radius of convergence, we can differentiate term by term and get that

$$\frac{d}{dt}P(t) = \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} Q^{k-1} \cdot Q = P(t)Q.$$

Obviously $P(0) = I$. Suppose that $\widetilde{P}$ is another solution. Then

$$\frac{d}{dt}(\widetilde{P}(t)e^{-tQ}) = \widetilde{P}'(t)e^{-tQ} + \widetilde{P}(t)\frac{d}{dt}(e^{-tQ}) = \widetilde{P}(t)Qe^{-tQ} - \widetilde{P}(t)Qe^{-tQ} = 0$$

and since $\widetilde{P}(0) = I$, we get $\widetilde{P}(t) = e^{tQ}$.

The case of the forward equations follows in exactly the same way.

(4) Differentiating $k$ times term by term gives the claimed identity. □

**Theorem 1.28.** *Let $S$ be a finite state space and let $Q$ be a matrix. Then it is a $Q$-matrix if and only if $P(t) = e^{tQ}$ is a stochastic matrix for all $t \geq 0$.*

**Proof.** For $t \downarrow 0$ we can write

$$P(t) = I + tQ + O(t^2).$$

Hence for all $x \neq y$ we get that $q_{xy} \geq 0$ if and only if $p_{xy}(t) \geq 0$ for all $t \geq 0$ sufficiently small. Since $P(t) = P(t/n)^n$ for all $n$ we obtain that $q_{xy} \geq 0$ for $x \neq y$ if and only if $p_{xy}(t) \geq 0$ for all $t \geq 0$ and all $x, y$.

Assume now that $Q$ is a $Q$-matrix. Then $\sum_y q_{xy} = 0$ for all $x$ and

$$\sum_y q_{xy}(n) = \sum_y \sum_z q_{xz}(n-1) q_{zy} = \sum_z q_{xz}(n-1) \sum_y q_{zy} = 0,$$

i.e. also $Q^n$ has zero sum rows. Hence we obtain

$$\sum_y p_{xy}(t) = 1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \sum_y q_{xy}(n) = 1,$$

which means that $P(t)$ is a stochastic matrix.

Assume now that $P(t)$ is a stochastic matrix. Then

$$\sum_y q_{xy} = \frac{d}{dt}\Big|_{t=0} \sum_y p_{xy}(t) = 0,$$

which shows that $Q$ is a $Q$-matrix. □

**Theorem 1.29.** *Let $X$ be a right continuous process with values in a finite set $S$ and let $Q$ be a $Q$-matrix on $S$. Then the following are equivalent:*

(a) *the process $X$ is Markov with generator $Q$;*

(b) *[infinitesimal definition] conditional on $X_s = x$ the process $(X_{s+t})_{t \geq 0}$ is independent of $(X_r)_{r \leq s}$ and uniformly in $t$ as $h \downarrow 0$ for all $x, y$*

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) = \mathbf{1}(x = y) + q_{xy} h + o(h);$$

(c) *for all $n \geq 0$, $0 \leq t_0 \leq \ldots \leq t_n$ and all states $x_0, \ldots, x_n$*

$$\mathbb{P}\big(X_{t_n} = x_n \bigm| X_{t_{n-1}} = x_{n-1}, \ldots, X_{t_0} = x_0\big) = p_{x_{n-1} x_n}(t_n - t_{n-1}),$$

*where $(p_{xy}(t))$ is the solution to the forward equation*

$$P'(t) = P(t)Q \quad \textit{and} \quad P(0) = I.$$

**Proof.** The equivalence between (a) and (c) follows from Theorem 1.23. We show the equivalence between (b) and (c).

(c)⇒(b) Since the state space is finite, from Theorem 1.27 we obtain that $P(t) = e^{tQ}$ and it is the unique solution to both the forward and backward equation. Then as $t \downarrow 0$ we get

$$P(t) = I + tQ + O(t^2).$$

So for all $t > 0$ as $h \downarrow 0$ for all $x, y$

$$\mathbb{P}(X_{t+h} = y \mid X_t = x) = \mathbf{1}(x = y) + q_{xy}h + o(h).$$

(b)⇒(c) For all $x, y \in S$ we have uniformly for all $t$ as $h \downarrow 0$

$$p_{xy}(t+h) = \sum_z \mathbb{P}_x(X_{t+h} = y, X_t = z) = \sum_z (\mathbf{1}(z = y) + q_{zy}h + o(h))\, p_{xz}(t),$$

and rearranging we get

$$\frac{p_{xy}(t+h) - p_{xy}(t)}{h} = \sum_z q_{zy}p_{xz}(t) + O(h),$$

since the state space is finite. Because this holds uniformly for all $t$, we can replace $t+h$ by $t-h$ as in the proof of Theorem 1.23 and thus like before we can differentiate and deduce

$$p'_{xy}(t) = \sum_z p_{xz}(t)q_{zy},$$

which shows that $(p_{xy}(t))$ satisfies the forward equation and this finishes the proof. □

We now end this section by giving an example of a $Q$-matrix on a state space with 3 elements and how we calculate $P(t) = e^{tQ}$. Suppose that

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

Then in order to calculate $P(t) = e^{tQ}$ we first diagonalise $Q$. The eigenvalues are $0, -2, -4$. Then we can write $Q = UDU^{-1}$ where $D$ is diagonal. In this case we obtain

$$e^{tQ} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!} = U \sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} 0^k & 0 & 0 \\ 0 & (-2t)^k & 0 \\ 0 & 0 & (-4t)^k \end{pmatrix} U^{-1} = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-2t} & 0 \\ 0 & 0 & e^{-4t} \end{pmatrix} U^{-1}.$$

Thus $p_{11}(t)$ must have the form $p_{11}(t) = a + be^{-2t} + ce^{-4t}$ for some constants $a, b, c$. We know $p_{11}(0) = 1$, $p'_{11}(0) = q_{11}$ and $p''_{11}(0) = q_{11}(2)$.

## 1.8 Non-minimal chains

So far we have only considered minimal chains, i.e. those that jump to a cemetery state after explosion. For such chains we established that the transition semigroup is the minimal non-negative solution to the forward and backward equations.

Minimal right-continuous processes are the simplest processes but also have a lot of applications. Let's see now what changes when we do not require the chain to die after explosion.

**Example 1.30.** Consider a birth process with rates $q_i = 2^i$. Then as we proved in Proposition 1.12 this process explodes almost surely. Suppose now that after explosion the chain goes back to 0 and restarts independently of what happened so far. After the second explosion it restarts again and so on. We denote this process by $\widetilde{X}$.

Then it clearly satisfies the Markov property, i.e.

$$\mathbb{P}\Big(\widetilde{X}_{t_{n+1}} = x_{n+1} \,\Big|\, \widetilde{X}_{t_n} = x_n, \ldots, \widetilde{X}_{t_0} = x_0\Big) = \widetilde{p}_{x_n x_{n+1}}(t_{n+1} - t_n)$$

and $\widetilde{p}_{xy}(t)$ also satisfies the backward and forward equations but it is not minimal. Indeed, if $X$ is a minimal birth chain with the same parameters, then

$$p_{xy}(t) = \mathbb{P}_x(X_t = y) = \mathbb{P}_x(X_t = y, t < \zeta),$$

while

$$\widetilde{p}_{xy}(t) = \mathbb{P}_x\Big(\widetilde{X} = y, t < \zeta\Big) + \mathbb{P}_x\Big(\widetilde{X}_t = y, t \ge \zeta\Big) = \mathbb{P}_x(X = y, t < \zeta) + \mathbb{P}_x\Big(\widetilde{X}_t = y, t \ge \zeta\Big).$$

Since $\mathbb{P}_x\Big(\widetilde{X}_t = y, t \ge \zeta\Big) > 0$, it immediately follows that $\widetilde{p}_{xy}(t) > p_{xy}(t)$.

In general in order to characterize a non-minimal chain we need in addition to the generator $Q$ also the way in which it restarts after explosion.

# 2 Qualitative properties of continuous time Markov chains

We first note that from now on we will only be dealing with minimal continuous time Markov chains, i.e. those that die after explosion. We will see that qualitative properties of the chain are the same as those for the jump chain. The state space will always be countable or finite.

## 2.1 Class structure

Let $x, y \in S$. We say that $x$ *leads* to $y$ and denote it by $x \to y$ if $\mathbb{P}_x(X_t = y \text{ for some } t \ge 0) > 0$. We say that $x$ and $y$ *communicate* if $x \to y$ and $y \to x$. The notions of communicating class, irreducibility, closed class and absorbing state are the same as in the discrete setting.

**Theorem 2.1.** *Let $X$ be a continuous time Markov chain with generator $Q$ and transition semi-group $(P(t))$. For any two states $x$ and $y$ the following are equivalent:*

(i) *$x \to y$;*

(ii) *$x \to y$ for the jump chain;*

(iii) *$q_{x_0x_1} \ldots q_{x_{n-1}x_n} > 0$ for some $x = x_0, \ldots, x_n = y$;*

(iv) *$p_{xy}(t) > 0$ for all $t > 0$;*

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

**Proof.** Implications (iv)⇒(v)⇒(i)⇒(ii) are immediate.

(ii)⇒(iii): Since $x \to y$ for the jump chain, it follows that there exist $x_0 = x, x_1, \ldots, x_n = y$ such that

$$p_{x_0x_1} \cdots p_{x_{n-1}x_n} > 0.$$

By the definition of the jump matrix, it follows that $q_{x_ix_{i+1}} > 0$ for all $i = 0, \ldots, n-1$, and this proves (iii).

(iii)⇒(iv): We first note that if for two states $w, z$ we have $q_{wz} > 0$, then

$$p_{wz}(t) \geq \mathbb{P}_w(J_1 \leq t, Y_1 = z, S_2 > t) = \left(1 - e^{-q_w t}\right) \frac{q_{wz}}{q_w} e^{-q_z t} > 0.$$

Hence assuming (iii) we obtain

$$p_{xy}(t) \geq p_{x_0 x_1}(t/n) \dots p_{x_{n-1} x_n}(t/n) > 0 \quad \forall t > 0,$$

where the inequality follows, since one way to go from $x$ to $y$ in $t$ steps is via the path $x_1, \dots, x_n$. □

## 2.2 Hitting times

Suppose that $X$ is a continuous time Markov chain with generator $Q$ on $S$. Let $Y$ be the jump chain and $A \subseteq S$. We set

$$T_A = \inf\{t > 0 : X_t \in A\} \quad \text{and} \quad H_A = \inf\{n \geq 0 : Y_n \in A\}.$$

Since $X$ is minimal, it follows that

$$\{T_A < \infty\} = \{H_A < \infty\} \quad \text{and on this event} \quad T_A = J_{H_A}.$$

Let $h_A(x) = \mathbb{P}_x(T_A < \infty)$ and $k_A(x) = \mathbb{E}_x[T_A]$.

**Theorem 2.2.** *Let $A \subseteq S$. The vector of hitting probabilities $(h_A(x))_{x \in S}$ is the minimal non-negative solution to the system of linear equations*

$$\begin{aligned} h_A(x) &= 1 \quad \forall\, x \in A \\ Qh_A(x) = \sum_{y \in S} q_{xy} h_A(y) &= 0 \quad \forall\, x \notin A. \end{aligned}$$

**Proof.** For the jump chain we have that $h_A(x)$ is the minimal non-negative solution to the following system of linear equations:

$$\begin{aligned} h_A(x) &= 1 \quad \forall\, x \in A \\ h_A(x) &= \sum_{y \neq x} h_A(y) p_{xy} \quad \forall\, x \notin A. \end{aligned}$$

The second equation can be rewritten

$$q_x h_A(x) = \sum_{y \neq x} h_A(y) q_{xy} \Rightarrow \sum_y h_A(y) q_{xy} = 0 \Rightarrow Qh(x) = 0.$$

The proof of minimality follows from the discrete case. □

**Theorem 2.3.** *Let $X$ be a continuous time Markov chain with generator $Q$ and $A \subseteq S$. Suppose that $q_x > 0$ for all $x \notin A$. Then $k_A(x) = \mathbb{E}_x[T_A]$ is the minimal non-negative solution to*

$$\begin{aligned} k_A(x) &= 0 \quad \forall\, x \in A \\ Qk_A(x) = \sum_y q_{xy} k_A(y) &= -1 \quad \forall\, x \notin A. \end{aligned}$$

**Proof.** Clearly if $x \in A$, then $T_A = 0$, and hence $k_A(x) = 0$. Let $x \notin A$. Then $T_A \geq J_1$ and by the Markov property we obtain

$$\mathbb{E}_x[T_A - J_1 \mid Y_1 = y] = \mathbb{E}_y[T_A]\,.$$

By conditioning on $Y_1$ we thus get

$$k_A(x) = \mathbb{E}_x[J_1] + \sum_{y \neq x} p_{xy}\mathbb{E}_y[T_A] = \frac{1}{q_x} + \sum_{y \neq x} \frac{q_{xy}}{q_x} k_A(y).$$

Therefore for $x \notin A$ we showed $Qk_A(x) = -1$. The proof of minimality follows in the same way as in the discrete case and is thus omitted. □

**Remark 2.4.** We note that the hitting probabilities are the same for the jump chain and the continuous time Markov chain. However, the expected hitting times differ, since in the continuous case we have to take into account also the exponential amount of time that the chain spends at each state.

## 2.3 Recurrence and transience

**Definition 2.5.** We call a state $x$ *recurrent* if $\mathbb{P}_x(\{t : X_t = x\} \text{ is unbounded}) = 1$. We call $x$ *transient* if $\mathbb{P}_x(\{t : X_t = x\} \text{ is unbounded}) = 0$.

**Remark 2.6.** We note that if $X$ explodes with positive probability starting from $x$, i.e. if $\mathbb{P}_x(\zeta < \infty) > 0$, then $x$ cannot be recurrent.

As in the discrete setting we have the following dichotomy.

**Theorem 2.7.** *Let $X$ be a continuous time Markov chain and $Y$ its jump chain. Then*

(i) *If $x$ is recurrent for $Y$, then $x$ is recurrent for $X$.*

(ii) *If $x$ is transient for $Y$, then $x$ is transient for $X$.*

(iii) *Every state is either recurrent or transient.*

(iv) *Recurrence and transience are class properties.*

**Proof.** (i) Suppose that $x$ is recurrent for $Y$ and $X_0 = x$. Then $X$ cannot explode, and hence $\mathbb{P}_x(\zeta = \infty) = 1$, or equivalently $J_n \to \infty$ with probability 1 starting from $x$. Since $X_{J_n} = Y_n$ for all $n$ and $Y$ visits $x$ infinitely many times, it follows that the set $\{t \geq 0 : X_t = x\}$ is unbounded.

(ii) If $x$ is transient, then $q_x > 0$, otherwise $x$ would be absorbing for $Y$. Hence

$$N = \sup\{n : Y_n = x\} < \infty$$

and if $t \in \{s : X_s = x\}$, then $t \leq J_{N+1}$. Since $(Y_n : n \leq N)$ cannot contain any absorbing state, it follows that $J_{N+1} < \infty$, and therefore $x$ is transient for $X$.

(iii), (iv) Recurrence and transience are class properties in the discrete setting, hence from (i) and (ii) we deduce the same for the continuous setting. □

**Theorem 2.8.** *The state $x$ is recurrent for $X$ if and only if $\int_0^\infty p_{xx}(t)\,dt = \infty$. The state $x$ is transient for $X$ if and only if $\int_0^\infty p_{xx}(t)\,dt < \infty$.*

**Proof.** If $q_x = 0$, then $x$ is recurrent and $p_{xx}(t) = 1$ for all $t$, and hence $\int_0^\infty p_{xx}(t)\,dt = \infty$.

Suppose now that $q_x > 0$. Then it suffices to show that

$$\int_0^\infty p_{xx}(t)\,dt = \frac{1}{q_x} \cdot \sum_{n=0}^\infty p_{xx}(n), \tag{2.1}$$

since then the result will follow from the dichotomy in the discrete time setting. We now turn to prove (2.1). We have

$$\int_0^\infty p_{xx}(t)\,dt = \mathbb{E}_x\left[\int_0^\infty \mathbf{1}(X_t = x)\,dt\right] = \mathbb{E}_x\left[\sum_{n=0}^\infty S_{n+1}\mathbf{1}(Y_n = x)\right] = \sum_{n=0}^\infty \mathbb{E}_x[S_{n+1}\mathbf{1}(Y_n = x)]$$
$$= \sum_{n=0}^\infty \mathbb{P}_x(Y_n = x)\,\mathbb{E}_x[S_{n+1} \mid Y_n = x] = \frac{1}{q_x} \cdot \sum_{n=0}^\infty p_{xx}(n),$$

where the third equality follows from Fubini and the last one from the fact that conditional on $Y_n = x$ the holding time $S_{n+1}$ is Exponentially distributed with parameter $q_x$. □

## 2.4 Invariant distributions

**Definition 2.9.** Let $Q$ be the generator of a continuous time Markov chain and let $\lambda$ be a measure. It is called *invariant* if $\lambda Q = 0$.

As in discrete time, also here the term invariant means that if we start the chain according to this distribution, then at each time $t$ the distribution will be the same. We will prove this later.

For $x \in S$ we let $T_x = \inf\{t \geq J_1 : X_t = x\}$ and $H_x = \inf\{n \geq 1 : Y_n = x\}$ be the first return times to $x$ for $X$ and $Y$ respectively.

First we show how invariant measures of the jump chain and of the continuous time chain are related.

**Theorem 2.10.** *Let $X$ be a continuous time chain with generator $Q$ and let $Y$ be its jump chain. The measure $\pi$ is invariant for $X$ if and only if the measure $\mu$ defined by $\mu_x = q_x \pi_x$ is invariant for $Y$.*

**Proof.** First note that $q_x(p_{xy} - \mathbf{1}(x = y)) = q_{xy}$ and so we get

$$\sum_x \pi(x) q_{xy} = \sum_x \pi(x) q_x (p_{xy} - \mathbf{1}(x = y)) = \sum_x \mu(x)(p_{xy} - \mathbf{1}(x = y)).$$

Therefore, $\mu P = \mu \Leftrightarrow \pi Q = 0$. □

We now recall a result from discrete time theory concerning invariant measures.

**Theorem 2.11.** *Let $Y$ be a discrete time Markov chain which is irreducible and recurrent and let $x$ be a state. Then the measure*

$$\nu(y) = \mathbb{E}_x\left[\sum_{n=0}^{H_x - 1} \mathbf{1}(Y_n = y)\right]$$

*is invariant for $Y$ and for all $y$ it satisfies $0 < \nu(y) < \infty$ and $\nu(x) = 1$.*

**Theorem 2.12.** *Let $Y$ be an irreducible discrete time Markov chain and $\nu$ as in Theorem 2.11. If $\lambda$ is another invariant measure with $\lambda(x) = 1$, then $\lambda(y) \geq \nu(y)$ for all $y$. If $Y$ is recurrent, then $\lambda(y) = \nu(y)$ for all $y$.*

**Theorem 2.13.** *Let $X$ be an irreducible and recurrent continuous time Markov chain with generator $Q$. Then $X$ has an invariant measure which is unique up to scalar multiplication.*

**Proof.** Since the chain is irreducible, it follows that $q_x > 0$ for all $x$ (except if the state space is a singleton, which is a trivial case).

The jump chain $Y$ will then also be irreducible and recurrent, and hence the measure $\nu$ defined in Theorem 2.11 is invariant for $Y$ and it is unique up to scalar multiplication.

Since we assumed that $q_x > 0$, from Theorem 2.10 we get that the measure $\pi(x) = \nu(x)/q_x$ is invariant for $X$ and it is unique up to scalar multiples. □

Just like in the discrete time setting, also here the existence of a stationary probability distribution is related to the question of positive recurrence.

Recall $T_x = \inf\{t \geq J_1 : X_t = x\}$ is the first return time to $x$.

**Definition 2.14.** A recurrent state $x$ is called *positive recurrent* if $m_x = \mathbb{E}_x[T_x] < \infty$. Otherwise it is called *null recurrent.*

**Theorem 2.15.** *Let $X$ be an irreducible continuous time Markov chain with generator $Q$. The following statements are equivalent:*

(i) *every state is positive recurrent;*

(ii) *some state is positive recurrent;*

(iii) *$X$ is non-explosive and has an invariant distribution $\lambda$.*

*Furthermore, when* (iii) *holds, then $\lambda(x) = (q_x m_x)^{-1}$ for all $x$.*

**Proof.** Again we assume that $q_x > 0$ for all $x$ by irreducibility, since otherwise the result is trivial.

The implication (i)$\Longrightarrow$(ii) is obvious. We show first that (ii)$\Longrightarrow$(iii). Let $x$ be a positive recurrent state. Then it follows that all states have to be recurrent, and hence $X$ is non-explosive, i.e. $\mathbb{P}_y(\zeta = \infty) = 1$ for all $y \in S$. For all $y \in S$ we define the measure

$$\mu(y) = \mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}(X_s = y)\, ds\right], \tag{2.2}$$

i.e. $\mu(y)$ is the expected amount of time the chain spends at $y$ in an excursion from $x$. Recall that $H_x$ denotes the first return time to $x$ for the jump chain. We can rewrite $\mu(y)$ using the jump chain $Y$ and holding times $(S_n)$ as follows

$$\begin{aligned}\mu(y) &= \mathbb{E}_x\left[\sum_{n=0}^{\infty} S_{n+1}\mathbf{1}(Y_n = y, n < H_x)\right] = \sum_{n=0}^{\infty} \mathbb{E}_x[S_{n+1} \mid Y_n = y, n < H_x]\, \mathbb{P}_x(Y_n = y, n < H_x) \\ &= \frac{1}{q_y} \cdot \mathbb{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}(Y_n = y, n < H_x)\right] = \frac{1}{q_y} \cdot \mathbb{E}_x\left[\sum_{n=0}^{H_x - 1} \mathbf{1}(Y_n = y)\right] = \frac{\nu(y)}{q_y},\end{aligned}$$

where the second equality follows from Fubini's theorem. Since the jump chain is recurrent, it follows that $\nu$ is an invariant measure for $Y$. Using Theorem 2.10 it follows that $\mu$ is invariant for $X$. Taking the sum over all $y$ of $\mu(y)$ we obtain

$$\sum_y \mu(y) = \mathbb{E}_x[T_x] = m_x < \infty,$$

and hence we can normalise $\mu$ in order to get a probability distribution, i.e.

$$\lambda(y) = \frac{\mu(y)}{m_x}$$

is an invariant distribution, which satisfies $\lambda(x) = (q_x m_x)^{-1}$.

Suppose now that (iii) holds. By Theorem 2.10 the measure $\beta(y) = \lambda(y)q_y$ is invariant for $Y$. Since $Y$ is irreducible and $\sum_y \lambda(y) = 1$, we have that $q_y\lambda(y) \geq q_x\lambda(x)p_{xy}(n) > 0$ for some $n > 0$. Hence $\lambda(y) > 0$ for all $y$ and since we have assumed that $q_y > 0$ for all $y$, we can define the measure $a(y) = \beta(y)/(\lambda(x)q_x)$. This is invariant for the jump chain and satisfies $a(x) = 1$. From Theorem 2.12 we now obtain that for all $y$

$$a(y) \geq \mathbb{E}_x\left[\sum_{n=0}^{H_x-1} \mathbf{1}(Y_n = y)\right].$$

We also have $\mu(y) = \nu(y)/q_y$ and since $X$ is non-explosive, we have $m_x = \sum_y \mu(y)$. Thus we deduce

$$m_x = \sum_y \mu(y) = \sum_y \frac{1}{q_y}\mathbb{E}_x\left[\sum_{n=0}^{H_x-1} \mathbf{1}(Y_n = y)\right] \leq \sum_y \frac{a(y)}{q_y} = \frac{1}{\lambda(x)q_x} \cdot \sum_y \lambda(y) = \frac{1}{\lambda(x)q_x} < \infty. \tag{2.3}$$

Therefore all states are positive recurrent.

Note that we did not need to use the relationship between $m_x$ and $\lambda(x)$. Hence if (iii) holds, i.e. if there is an invariant distribution and the chain does not explode, then this implies that all states are positive recurrent. Therefore the jump chain is recurrent and invoking Theorem 2.12 we get that $a(y) = \nu(y)$ for all $y$. (Note that we get the same equality for all starting points $x$.) Hence the inequality in (2.3) becomes an equality and this proves that $\lambda(x) = (q_x m_x)^{-1}$ for all states $x$. This concludes the proof. □

**Example 2.16.** Consider a birth and death chain $X$ on $\mathbb{Z}_+$ with transition rates $q_{i,i+1} = \lambda q_i$ and $q_{i,i-1} = \mu q_i$ with $q_i > 0$ for all $i$. It is not hard to check that the measure $\pi_i = q_i^{-1}\left(\frac{\lambda}{\mu}\right)^i$ is invariant. (We will see later how one can find this invariant measure by solving the detailed balance equations which is equivalent to reversibility.) Taking $q_i = 2^i$ and $\lambda = 3\mu/2$ we see that $\pi$ can be normalized to give an invariant distribution. However, when $\lambda > \mu$, then the jump chain is transient, and hence also $X$ is transient.

Therefore we see that in the continuous time setting the existence of an invariant distribution is not equivalent to positive recurrence if the chain explodes as in this example.

We now explain the terminology *invariant measure*. First we treat the finite state space case and then move to the countable one.

**Theorem 2.17.** *Let $X$ be a continuous time Markov chain with generator $Q$ on a finite state space $S$. Then $\pi P(t) = \pi \ \forall t \geq 0 \Leftrightarrow \pi$ is invariant.*

**Proof.** The transition semigroup satisfies $P'(t) = P(t)Q = QP(t)$. Differentiating $\pi P(t) = \pi$ with respect to $t$ we get

$$\frac{d}{dt}(\pi P(t)) = \pi P'(t) = \pi Q P(t) = \pi P(t) Q,$$

where in the first equality we were able to interchange differentiation and sum, because $|S| < \infty$. Therefore if $\pi P(t) = \pi$ for all $t$, then from the above equality we obtain

$$0 = \frac{d}{dt}(\pi P(t)) = \pi P(t) Q = \pi Q.$$

If $\pi Q = 0$, then $\frac{d}{dt}(\pi P(t)) = 0$ and thus $\pi P(t) = \pi P(0) = \pi$ for all $t$. □

Before proving the general countable state space case we state and prove an easy lemma.

**Lemma 2.18.** *Let $X$ be a continuous time Markov chain and let $t > 0$ be fixed. Set $Z_n = X_{nt}$.*

(i) *If $x$ is recurrent for $X$, then $x$ is recurrent for $Z$ too.*

(ii) *If $x$ is transient for $X$, then $x$ is transient for $Z$ too.*

**Proof.** Part (ii) is obvious, since if the set of times that $X$ visits $x$ is bounded, then $Z$ cannot visit $x$ infinitely many times.

Suppose now that $x$ is recurrent for $X$. We will use the criterion for recurrence in terms of the convergence of the sum $\sum_{n=0}^{\infty} p_{xx}(n)$. We divide time into intervals of length $t$. Then we have

$$\int_0^\infty p_{xx}(s)\, ds = \sum_{n=0}^{\infty} \int_{tn}^{t(n+1)} p_{xx}(s)\, ds.$$

For $s \in (tn, t(n+1))$ by the Markov property we have $p_{xx}((n+1)t) \geq e^{-q_x t} p_{xx}(s)$. Therefore, we deduce

$$\int_0^\infty p_{xx}(s)\, ds \leq t e^{q_x t} \sum_{n=0}^{\infty} p_{xx}((n+1)t).$$

Since $x$ is recurrent for $X$, by Theorem 2.8 we obtain that the integral on the left hand side above is infinite, and hence

$$\sum_{n=0}^{\infty} p_{xx}(nt) = \infty$$

and this finishes the proof. □

**Theorem 2.19.** *Let $X$ be a recurrent continuous time Markov chain on a countable state space $S$ with generator $Q$ and let $\lambda$ be a measure. Then $\lambda Q = 0 \Leftrightarrow \lambda P(s) = \lambda \ \forall s > 0$.*

**Proof.** We start by showing that any measure $\lambda$ satisfying $\lambda Q = 0$ or $\lambda P(s) = \lambda$ for all $s > 0$ is unique up to scalar multiples. Indeed, if $\lambda Q = 0$, then this follows from Theorem 2.13. Suppose now that $\lambda P(t) = \lambda$ and consider the discrete time Markov chain with transition matrix $P(t)$. Then this is clearly irreducible (because $X$ is irreducible and $p_{xy}(t) > 0$ for all $x, y$ from Theorem 2.1) and recurrent from Lemma 2.18, and $\lambda$ is an invariant measure. This now implies that also in this case $\lambda$ is unique up to scalar multiples.

We showed in the proof of Theorem 2.15 that the measure

$$\mu(y) = \mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}(X_s = y)\, ds\right] = \frac{\nu(y)}{q_y}$$

is invariant, i.e. $\mu Q = 0$. (Note that since the discrete time chain is recurrent, we have $\nu(y) < \infty$ for all $y$ by Theorem 2.11, and hence the multiplication $\mu Q$ makes sense.)

To finish the proof we now need to show that $\mu P(s) = \mu$ for all $s > 0$. By the strong Markov property for the jump chain we get

$$\mathbb{E}_x\left[\int_0^{s} \mathbf{1}(X_t = y)\, dt\right] = \mathbb{E}_x\left[\int_{T_x}^{T_x+s} \mathbf{1}(X_t = y)\, dt\right]. \tag{2.4}$$

Since we can write

$$\int_0^{T_x} \mathbf{1}(X_t = y)\, dt = \int_0^{s} \mathbf{1}(X_t = y)\, dt + \int_s^{T_x} \mathbf{1}(X_t = y)\, dt,$$

using (2.4) and Fubini's theorem we obtain

$$\begin{aligned}
\mu(y) &= \mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}(X_t = y)\, dt\right] = \mathbb{E}_x\left[\int_0^{s} \mathbf{1}(X_t = y)\, dt\right] + \mathbb{E}_x\left[\int_s^{T_x} \mathbf{1}(X_t = y)\, dt\right] \\
&= \mathbb{E}_x\left[\int_{T_x}^{T_x+s} \mathbf{1}(X_t = y)\, dt\right] + \mathbb{E}_x\left[\int_s^{T_x} \mathbf{1}(X_t = y)\, dt\right] = \mathbb{E}_x\left[\int_s^{s+T_x} \mathbf{1}(X_t = y)\, dt\right] \\
&= \int_0^\infty \sum_{z\in S} \mathbb{P}_x(X_t = z, X_{t+s} = y, t < T_x)\ dt = \sum_{z\in S} p_{zy}(s)\mathbb{E}_x\left[\int_0^{T_x} \mathbf{1}(X_t = z)\, dt\right] \\
&= \sum_{z\in S} \mu(z) p_{zy}(s),
\end{aligned}$$

which shows that $\mu P(s) = \mu$ and this completes the proof. □

**Remark 2.20.** Note that in the above theorem we required $X$ to be recurrent in order to prove that the equivalence holds. We now explain that if we only assume that $\pi P(t) = \pi$ for all $t$ and the chain is irreducible, then $\pi Q = 0$.

Consider the discrete time Markov chain $Z$ with transition matrix $P(t)$ (this is always stochastic, since $\pi P(t) = \pi$ and $\pi$ is a distribution). Then $Z$ is clearly irreducible and has an invariant distribution. Therefore it is positive recurrent, and Lemma 2.18 implies that $X$ is also recurrent. Then applying the above theorem we get that $\pi Q = 0$.

## 2.5 Convergence to equilibrium

An irreducible, aperiodic and positive recurrent discrete time Markov chain converges to equilibrium as time goes to $\infty$. The same is true in the continuous setting, but we no longer need to assume the chain is aperiodic, since in continuous time as we showed in Theorem 2.1 $p_{xy}(t) > 0$ for all $t > 0$.

**Lemma 2.21.** *Let $Q$ be a $Q$-matrix with semigroup $P(t)$. Then*

$$|p_{xy}(t+h) - p_{xy}(t)| \leq 1 - e^{-q_x h}.$$

**Proof.** By Chapman Kolmogorov we have

$$\begin{aligned}
|p_{xy}(t+h) - p_{xy}(t)| &= \left|\sum_z p_{xz}(h)p_{zy}(t) - p_{xy}(t)\right| = \left|\sum_{z\neq x} p_{xz}(h)p_{zy}(t) + p_{xx}(h)p_{xy}(t) - p_{xy}(t)\right| \\
&= \left|\sum_{z\neq x} p_{xz}(h)p_{zy}(t) - p_{xy}(t)(1-p_{xx}(h))\right| \\
&\leq 1 - p_{xx}(h) \leq \mathbb{P}_x(J_1 \leq h) = 1 - e^{-q_x h}
\end{aligned}$$

and this finishes the proof. □

**Theorem 2.22** (Convergence to equilibrium)**.** *Let $X$ be an irreducible non explosive continuous time chain on $S$ with generator $Q$. Suppose that $\lambda$ is an invariant distribution. Then for all $x, y \in S$ we have*

$$p_{xy}(t) \to \lambda(y) \quad \textit{as} \quad t \to \infty.$$

**Proof.** Fix $h > 0$ such that $1 - e^{-q_x h} \leq \varepsilon/2$ and let $Z_n = X_{nh}$ be a discrete time chain with transition matrix $P(h)$. Since $X$ is irreducible, Theorem 2.1 gives that $p_{xy}(h) > 0$ for all $x, y$ which shows that $Z$ is irreducible and aperiodic.

Since $X$ is irreducible, non-explosive and has an invariant distribution, Theorem 2.15 gives that $X$ is positive recurrent. Thus we can apply Theorem 2.19 to get that $\lambda$ is an invariant measure for $Z$. Hence applying the convergence to equilibrium theorem for discrete time chains we obtain for all $x, y$

$$p_{xy}(nh) \to \lambda(y) \quad \text{as} \quad n \to \infty.$$

This means that for all $\varepsilon > 0$, there exists $n_0$ such that for all $n \geq n_0$ we have

$$|p_{xy}(nh) - \lambda(y)| \leq \frac{\varepsilon}{2}.$$

Let $t \geq n_0 h$. Then there must exist $n \in \mathbb{N}$ with $n \geq n_0$ such that $nh \leq t < (n+1)h$. Therefore by the choice of $h$ we deduce

$$|p_{xy}(t) - \lambda(y)| \leq |p_{xy}(t) - p_{xy}(nh)| + |p_{xy}(nh) - \lambda(y)| \leq 1 - e^{-q_x(t-nh)} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Hence this shows that $p_{xy}(t) \to \lambda(y)$ as $t \to \infty$. □

## 2.6 Reversibility

**Theorem 2.23.** *Let $X$ be an irreducible and non-explosive continuous time Markov chain on $S$ with generator $Q$ and invariant distribution $\pi$. Suppose that $X_0 \sim \pi$. Fix $T > 0$ and set $\widehat{X}_t = X_{T-t}$ for $0 \leq t \leq T$. Then $\widehat{X}$ is Markov with generator $\widehat{Q}$ and invariant distribution $\pi$, where $\widehat{q}_{xy} = \pi(y)q_{yx}/\pi(x)$. Moreover, $\widehat{Q}$ is irreducible and non-explosive.*

**Proof.** We first note that $\widehat{Q}$ is a $Q$-matrix, since $\widehat{q}_{xy} \geq 0$ for all $x \neq y$ and

$$\sum_y \widehat{q}_{xy} = \sum_y \frac{\pi(y)}{\pi(x)} q_{yx} = \frac{1}{\pi(x)} \cdot (\pi Q)_x = 0,$$

since $\pi$ is invariant for $Q$.

Let $P(t)$ be the transition semigroup of $X$. Then $P(t)$ is the minimal non-negative solution of the forward equations, i.e. $P'(t) = P(t)Q$ and $P(0) = I$.

Let $0 = t_0 \leq t_1 \leq t_2 \leq \ldots t_n = T$ and $x_1, x_2, \ldots, x_n \in S$. Setting $s_i = t_i - t_{i-1}$, we have

$$\mathbb{P}\left(\widehat{X}_{t_0} = x_0, \ldots, \widehat{X}_{t_n} = x_n\right) = \mathbb{P}(X_T = x_0, \ldots, X_0 = x_n) = \pi(x_n) p_{x_n x_{n-1}}(s_n) \ldots p_{x_1 x_0}(s_1).$$

We now define

$$\widehat{p}_{xy}(t) = \frac{\pi(y)}{\pi(x)} p_{yx}(t).$$

Thus we obtain

$$\mathbb{P}\left(\widehat{X}_{t_0} = x_0, \ldots, \widehat{X}_{t_n} = x_n\right) = \pi(x_0) \widehat{p}_{x_0 x_1}(s_1) \ldots \widehat{p}_{x_{n-1} x_n}(s_n).$$

We now need to show that $\widehat{P}(t)$ is the minimal non-negative solution to Kolmogorov's backward equations with generator $\widehat{Q}$. Indeed, we have

$$\begin{aligned}\widehat{p}'_{xy}(t) &= \frac{\pi(y)}{\pi(x)} p'_{xy}(t) = \frac{\pi(y)}{\pi(x)} \cdot \sum_z p_{yz}(t) q_{zx} = \frac{1}{\pi(x)} \cdot \sum_z \pi(z) \widehat{p}_{zy}(t) q_{zx} \\ &= \sum_z \widehat{p}_{zy}(t) \cdot \frac{\pi(z)}{\pi(x)} q_{zx} = \sum_z \widehat{p}_{zy}(t) \widehat{q}_{xz} = (\widehat{Q}\widehat{P}(t))_{xy}.\end{aligned}$$

Next we show that $\widehat{P}$ is the minimal solution to these equations. Suppose that $R$ is another solution to these equations and define $\overline{R}_{xy}(t) = \frac{\pi(y)}{\pi(x)} R_{yx}(t)$. Then

$$\overline{R}'_{xy}(t) = \frac{\pi(y)}{\pi(x)} R'_{yx}(t) = \frac{\pi(y)}{\pi(x)} \sum_z \widehat{q}_{yz} R_{zx}(t) = \frac{\pi(y)}{\pi(x)} \sum_z q_{zy} \frac{\pi(z)}{\pi(y)} R_{zx}(t) = \sum_z \overline{R}_{zx}(t) q_{zy}.$$

Thus $\overline{R}$ satisfies Kolmogorov's forward equations, and hence $\overline{R} \geq P$, which implies that $R \geq \widehat{P}$.

Since $Q$ is irreducible and using the definition of $\widehat{Q}$ we deduce that $\widehat{Q}$ is also irreducible. Moreover, since $\pi$ is invariant for $Q$, we deduce

$$\sum_y \pi(y) \widehat{q}_{yx} = \sum_y \pi(y) \frac{\pi(x)}{\pi(y)} q_{xy} = 0,$$

which shows that $\pi$ is the invariant distribution of the Markov chain with generator $\widehat{Q}$.

It only remains to show that $\widehat{Q}$ does not explode. Let $\widehat{\zeta}$ be the explosion time of a Markov chain $Z$ with generator $\widehat{Q}$. Then

$$\widehat{p}_{xy}(t) = \mathbb{P}_x\left(Z_t = y, t < \widehat{\zeta}\right).$$

But by the definition of $\widehat{p}$ we have that $\sum_y \widehat{p}_{xy}(t) = 1$ (because $Q$ does not explode) and therefore $\mathbb{P}_x\left(t < \widehat{\zeta}\right) = 1$ for all $t$, which implies that $\widehat{\zeta} = \infty$ almost surely. □

In the same way as in discrete time, we define the notion of reversibility in the continuous setting.

**Definition 2.24.** Let $X$ be a Markov chain with generator $Q$. It is called *reversible* if for all $T > 0$ the processes $(X_t)_{0\leq t\leq T}$ and $(X_{T-t})_{0\leq t\leq T}$ have the same distribution.

A measure $\lambda$ and a $Q$-matrix $Q$ are said to be in *detailed balance* if for all $x, y$ we have

$$\lambda(x)q_{xy} = \lambda(y)q_{yx}.$$

**Lemma 2.25.** *Suppose that $Q$ and $\lambda$ are in detailed balance. Then $\lambda$ is an invariant distribution for $Q$.*

**Proof.** We need to show that $\lambda Q = 0$. Taking the sum

$$\sum_x \lambda(x)q_{xy} = \sum_x \lambda(y)q_{yx} = 0,$$

since $Q$ is a $Q$-matrix and this completes the proof. $\square$

**Theorem 2.26.** *Let $X$ be irreducible and non explosive with generator $Q$ and let $\pi$ be a probability distribution with $X_0 \sim \pi$. Then $\pi$ and $Q$ are in detailed balance $\Leftrightarrow$ $(X_t)_{t\geq 0}$ is reversible.*

**Proof.** If $\pi$ and $Q$ are in detailed balance, then $\widehat{Q} = Q$, where $\widehat{Q}$ is the matrix defined in Theorem 2.23 and $\pi$ is the invariant distribution of $\widehat{Q}$. Therefore, from Theorem 2.23 again, the reversed chain has the same distribution as $X$, and hence $X$ is reversible.

Suppose now that $X$ is reversible. Then $\pi$ is an invariant distribution. Then from Theorem 2.23 we get that $\widehat{Q} = Q$, which immediately gives that $\pi$ and $Q$ are in detailed balance. $\square$

**Remark 2.27.** As in discrete time, when we look for an invariant distribution, it is always easier to look for a solution to the detailed balance equations first. If there is no such solution, which means the chain is not reversible, then we have to do the matrix multiplication or use different methods.

**Definition 2.28.** A birth and death chain $X$ is a continuous time Markov chain on $\mathbb{N} = \{0, 1, 2, \ldots\}$ with non-zero transition rates $q_{x,x-1} = \mu_x$ and $q_{x,x+1} = \lambda_x$ for $x \in \mathbb{N}$.

**Lemma 2.29.** *A measure $\pi$ is invariant for a birth and death chain if and only if it solves the detailed balance equations.*

**Proof.** If $\pi$ solves the detailed balance equations, then $\pi$ is invariant by Lemma 2.25.

Suppose that $\pi$ is an invariant measure for $X$. Then $\pi Q = 0$ or equivalently $\sum_i \pi_i q_{i,j} = 0$ for all $j$. If $j \geq 1$ this means

$$\pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} = \pi_j(\lambda_j + \mu_j) \Leftrightarrow \pi_{j+1}\mu_{j+1} - \pi_j\lambda_j = \pi_j\mu_j - \pi_{j-1}\lambda_{j-1}.$$

For $j = 0$ we have $\pi_0\lambda_0 = \pi_1\mu_1$. Plugging this into the right hand side of the above equation and applying induction gives that for all $j$ we have

$$\pi_j\lambda_j = \pi_{j+1}\mu_{j+1},$$

i.e. the detailed balance equations hold. $\square$

## 2.7 Ergodic theorem

As in discrete time the long run proportion of time that a chain spends at a state $x$ is given by $1/\mathbb{E}_x[T_x^+]$ which is equal to the invariant probability of the state (if the chain is positive recurrent), the same is true in continuous time. The proof of the ergodic theorem in the continuous setting is similar to the discrete one with the only difference being that every time we visit the state we also spend an exponential amount of time there.

**Theorem 2.30.** *Let $Q$ be an irreducible $Q$-matrix and let $X$ be a Markov chain with generator $Q$ started from initial distribution $\nu$. Then almost surely we have*

$$\frac{1}{t}\int_0^t \mathbf{1}(X_s = x)\,ds \to \frac{1}{q_x m_x} \quad \text{as } t \to \infty,$$

*where $m_x = \mathbb{E}_x[T_x]$ and $T_x = \inf\{t \geq J_1 : X_t = x\}$. Moreover, in the positive recurrent case, if $f : S \to \mathbb{R}$ is a bounded function, then almost surely*

$$\frac{1}{t}\int_0^t f(X_s)\,ds \to \overline{f},$$

*where $\overline{f} = \sum_x f(x)\pi(x)$ and $\pi$ is the unique invariant distribution.*

**Proof.** First we note that if $Q$ is transient, then the set of times that $X$ visits $x$ is bounded, and hence almost surely

$$\frac{1}{t}\int_0^t \mathbf{1}(X_s = x)\,ds \to 0 \quad \text{as } t \to \infty$$

and also $m_x = \infty$.

Suppose now that $Q$ is recurrent. First let's see how we can obtain the result heuristically. Let $N(t)$ be the number of visits to $x$ up to time $t$. Then at every visit the Markov chain spends an exponential amount of time of parameter $q_x$. Hence

$$\frac{1}{t}\int_0^t \mathbf{1}(X_s = x)\,ds \approx \frac{\sum_{i=1}^{N(t)} S_i}{N(t)} \cdot \frac{N(t)}{t}.$$

By the law of large numbers we have that as $t \to \infty$

$$\frac{\sum_{i=1}^{N(t)} S_i}{N(t)} \to \frac{1}{q_x},$$

since by recurrence $N(t) \to \infty$ as $t \to \infty$ almost surely. We also have that after every visit to $x$ it takes time $m_x$ on average to hit it again after leaving it, so $N(t)/t \to 1/m_x$ as $t \to \infty$. Let's prove it now rigorously.

First of all we explain that it suffices to prove the result when $\nu = \delta_x$. Indeed, since we assumed $Q$ to be recurrent, it follows that $x$ will be hit in finite time almost surely. Let $H_1$ be the first hitting time of $x$. Then

$$\frac{1}{t}\int_0^t \mathbf{1}(X_s = x)\,ds = \frac{1}{t}\int_{H_1}^t \mathbf{1}(X_s = x)\,ds$$

and since $H_1 < \infty$, taking the limit as $t \to \infty$ does not change the long run proportion of time we spend in $x$ when we count it after hitting $x$ for the first time. Therefore we consider the case when $\nu = \delta_x$ from now on.

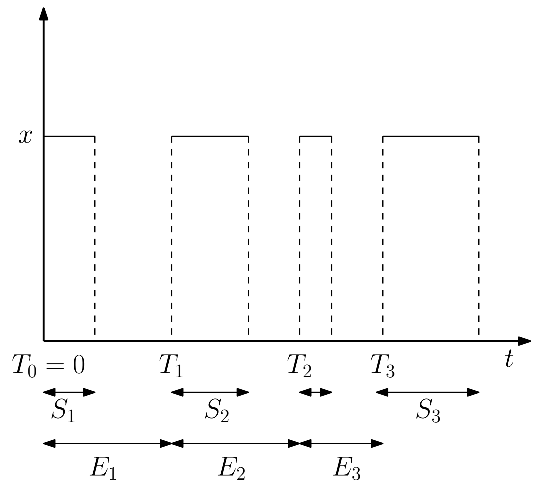



Figure 1: Visits to $x$

We now denote by $E_i$ the length of the $i$-th excursion from $x$, by $T_i$ the $i$-th visit to $x$ and by $S_i$ the amount of time spent at $x$ during the $i$-th visit as in Figure 1. More formally, we let $T_0 = 0$ and

$$\begin{aligned} S_{i+1} &= \inf\{t > T_i : X_t \neq x\} - T_i \\ T_{i+1} &= \inf\{t > T_i + S_{i+1} : X_t = x\} \\ E_{i+1} &= T_{i+1} - T_i. \end{aligned}$$

By the strong Markov property of the jump chain at the times $T_i$ we see that $(S_i)$ are i.i.d. exponentially distributed with parameter $q_x$ and also that $(E_i)$ are i.i.d. with mean $m_x$. By the strong law of large numbers we obtain that almost surely as $n \to \infty$

$$\frac{S_1 + \ldots + S_n}{n} \to \frac{1}{q_x} \quad \text{and} \quad \frac{E_1 + \ldots + E_n}{n} \to m_x. \tag{2.5}$$

Now for every $t$ we set $n(t)$ to be the smallest integer so that $T_{n(t)} \leq t < T_{n(t)+1}$. With this definition we get

$$S_1 + \ldots + S_{n(t)} \leq \int_0^t \mathbf{1}(X_s = x)\, ds < S_1 + \ldots + S_{n(t)+1}.$$

We also get that

$$E_1 + \ldots + E_{n(t)} \leq t < E_1 + \ldots + E_{n(t)+1}.$$

Therefore, dividing the above two inequalities we obtain

$$\frac{S_1 + \ldots + S_{n(t)}}{E_1 + \ldots + E_{n(t)+1}} \leq \frac{1}{t} \int_0^t \mathbf{1}(X_s = x)\, ds \leq \frac{S_1 + \ldots + S_{n(t)+1}}{E_1 + \ldots + E_{n(t)}}.$$

By recurrence we also see that $n(t) \to \infty$ as $t \to \infty$ almost surely. Hence using this together with the law of large numbers (2.5) we deduce that almost surely

$$\frac{1}{t} \int_0^t \mathbf{1}(X_s = x)\, ds \to \frac{1}{m_x q_x} \quad \text{as } t \to \infty.$$

We now turn to prove the statement of the theorem in the positive recurrent case, which implies that there is a unique invariant distribution $\pi$. We can write

$$\frac{1}{t}\int_0^t f(X_s)\,ds - \overline{f} = \sum_{x\in S} f(x)\left(\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right). \tag{2.6}$$

From the above we know that

$$\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x) \to 0 \quad \text{as } t \to \infty, \tag{2.7}$$

since $\pi(x) = (q_x m_x)^{-1}$. We next need to justify that the sum appearing in (2.6) also converges to 0. Suppose without loss of generality that $|f|$ is bounded by 1. The rest of the proof follows as in the discrete time case, but we include it here for completeness. Let $J$ be a finite subset of $S$. Then

$$\begin{aligned}
&\left|\sum_{x\in S} f(x)\left(\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right)\right| \le \sum_{x\in S}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| \\
&\le \sum_{x\in J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| + \sum_{x\notin J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| \\
&\le \sum_{x\in J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| + \sum_{x\notin J}\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) + \sum_{x\notin J}\pi(x) \\
&\le \sum_{x\in J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| + 1 - \sum_{x\in J}\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) + \sum_{x\notin J}\pi(x) \\
&\le 2\sum_{x\in J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| + 2\sum_{x\notin J}\pi(x).
\end{aligned}$$

We can choose $J$ finite so that $\sum_{x\notin J}\pi(x) < \varepsilon/4$. Using the finiteness of $J$ and (2.7) we get that there exists $t_0(\omega)$ sufficiently large so that

$$\sum_{x\in J}\left|\frac{1}{t}\int_0^t \mathbf{1}(X_s = x) - \pi(x)\right| \le \frac{\varepsilon}{4} \quad \text{for } t \ge t_0(\omega).$$

Therefore, taking $t \ge t_0(\omega)$ we conclude that

$$\left|\frac{1}{t}\int_0^t f(X_s)\,ds - \overline{f}\right| \le \varepsilon$$

and this finishes the proof. □

# 3 Queueing theory

## 3.1 Introduction

Suppose we have a succession of customers entering a queue, waiting for service. There are one or more servers delivering this service. Customers then move on, either leaving the system, or joining another queue, etc. The questions we have in mind are as follows: is there an equilibrium for the queue length? What is the expected length of the busy period (the time during which the server is busy serving customers until it empties out)? What is the total effective rate at which customers are being served? And how long do they spend in the system on average?

Queues form a convenient framework to address these and related issues. We will be using *Kendall's notation* throughout: e.g. the type of queue will be denoted by

$$M/G/k$$

- The first letter stands for the way customers arrive in the queue ($M$ = Markovian, i.e. a Poisson process with some rate $\lambda$).
- The second letter stands for the service time of customers ($G$ = general, i.e. no particular assumption is made on the distribution of the service time)
- The third letter stands for the number of servers in the system (typically $k = 1$ or $k = \infty$).

### 3.2 $M/M/1$ queue

Customers arrive according to a Poisson process of rate $\lambda > 0$. There is a single server and the service times are i.i.d. exponential with parameter $\mu > 0$. Let $X_t$ denote the queue length (including the customer being served at time $t \geq 0$). Then $X_t$ is a Markov chain on $S = \{0, 1, \ldots\}$ with

$$q_{i,i+1} = \lambda \quad \text{and} \quad q_{i,i-1} = \mu$$

and $q_{i,j} = 0$ if $j \neq i$ and $j \neq i \pm 1$. Hence $X$ is a birth and death chain.

**Theorem 3.1.** *Let $\rho = \lambda/\mu$. Then $X$ is transient if and only $\rho > 1$, recurrent if and only $\rho \leq 1$, and is positive recurrent if and only if $\rho < 1$. In the latter case $X$ has an equilibrium distribution given by*

$$\pi(n) = (1 - \rho)\rho^n.$$

*Suppose that $\rho < 1$, the queue is in equilibrium, i.e. $X_0 \sim \pi$ and $W$ is the waiting time of a customer that arrives at time $t$. Then the distribution of $W$ is* $\text{Exp}(\mu - \lambda)$.

**Proof.** The jump chain is given by a biased random walk on the integers with reflection at 0: the probability of jumping to the right is $p = \lambda/(\lambda + \mu)$. Hence the chain $X$ is transient if and only if $p > 1/2$ or equivalently $\lambda > \mu$, and recurrent otherwise. Concerning positive recurrence, observe that $\sup_i q_i < \infty$ so by Theorem 1.19 there is a.s. no explosion. Therefore by Theorem 2.15 positive recurrence is equivalent to the existence of an invariant distribution. Furthermore, since $X$ is a birth and death chain, by Lemma 2.29 it suffices to solve the Detailed Balance Equations, which read:

$$\pi(n)\lambda = \pi(n + 1)\mu$$

for all $n \geq 0$. We thus find $\pi(n + 1) = (\lambda/\mu)^{n+1}\pi(0)$ inductively and deduce the desired form for $\pi(n)$. Note that $\pi$ is the distribution of a (shifted) geometric random variable. (Shifted because it can be equal to 0).

Suppose now that $\rho < 1$ and $X_0 \sim \pi$. Suppose a customer arrives at time $t$ and let $N$ be the number of customers already in the queue at this time. Since $X_0 \sim \pi$, it follows that the distribution of $N$ is $\pi$. Then

$$W = \sum_{i=1}^{N+1} T_i,$$

where $(T_i)$ is an i.i.d. sequence of exponential random variables with parameter $\mu$ and independent of $N$. We also have that $N+1$ is a geometric variable that starts from 1. Therefore, using Exercise 2 from Example Sheet 1, we deduce that $W$ is exponential with parameter $\mu(1-\rho) = \mu - \lambda$. □

**Example 3.2.** What is the expected queue length at equilibrium? We have seen that the queue length $X$ at equilibrium is a shifted geometric random variable with success probability $1-\rho$. Hence

$$\mathbb{E}[X] = \frac{1}{1-\rho} - 1 = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}. \tag{3.1}$$

### 3.3 $M/M/\infty$ queue

Customers arrive at rate $\lambda$ and are served at rate $\mu$. There are infinitely many servers, so customers are in fact served immediately. Let $X_t$ denote the queue length at time $t$ (which consists only of customers being served at this time).

**Theorem 3.3.** *The queue length $X_t$ is a positive recurrent Markov chain for all $\lambda, \mu > 0$. Furthermore the invariant distribution is Poisson with parameter $\rho = \lambda/\mu$.*

**Proof.** The rates are $q_{i,i+1} = \lambda$ and $q_{i,i-1} = i\mu$ (since when there are $i$ customers in the queue, the first service will be completed after an exponential time with rate $i\mu$). Thus $X$ is a birth and death chain; hence for an invariant distribution it suffices to solve the Detailed Balance Equations:

$$\lambda \pi_{n-1} = n\mu\pi_n \; \forall\, n \Leftrightarrow \pi_n = \frac{1}{n}\frac{\lambda}{\mu}\pi_{n-1} = \ldots = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0.$$

Hence the Poisson distribution with parameter $\rho = \lambda/\mu$ is invariant. It remains to check that $X$ is not explosive. This is not straightforward as the rates are unbounded. However, we will show that the jump chain $Y$ is recurrent, and so using Theorem 1.19 this means that $X$ is non-explosive.

The jump chain is a birth and death chain on $\mathbb{N}$ with reflection at 0. The transition probabilities are

$$p_{i,i+1} = \frac{\lambda}{\mu i + \lambda} = 1 - p_{i,i-1}.$$

Let $k$ be sufficiently large so that $\mu n \geq 2\lambda$ for all $n \geq k$. Then this implies that for all $i \geq k$ we have

$$p_{i,i+1} \leq \frac{1}{3} \quad \text{and} \quad p_{i,i-1} \geq \frac{2}{3}. \tag{3.2}$$

Suppose we start $Y$ from $k+1$. We want to show that with probability 1 it will hit $k$ eventually. It follows that we can construct a $(2/3, 1/3)$ biased random walk $\widetilde{Y}_n$ such that $Y_n \leq \widetilde{Y}_n$ for all times $n$ up to the first time they hit $k$. But $\widetilde{Y}$ is transient towards $-\infty$ and hence is guaranteed to return to $k$ eventually.

Here is another way of proving that $Y$ is recurrent. We now set

$$\gamma_i = \frac{p_{i,i-1}p_{i-1,i-2}\cdots p_{1,0}}{p_{i,i+1}p_{i-1,i}\cdots p_{1,2}}.$$

From IB Markov chains (see for instance [2, page 16]) we know that if $\sum_i \gamma_i = \infty$, then the chain is recurrent. Indeed, in this case we have

$$\sum_i \gamma_i \geq A \sum_{i \geq k} 2^{i-k+2} = \infty,$$

and hence $Y$ is recurrent and therefore $X$ is non-explosive. This concludes the proof. □

## 3.4 Burke's theorem

Burke's theorem is one of the most intriguing (and beautiful) results of this course. Consider a $M/M/1$ queue and assume $\rho = \lambda/\mu < 1$, so there is an invariant distribution. Let $D_t$ denote the number of customers who have departed the queue up to time $t$.

**Theorem 3.4.** *(Burke's theorem). At equilibrium, $D_t$ is a Poisson process with rate $\lambda$, independently of $\mu$ (so long as $\mu > \lambda$). Furthermore, $X_t$ is independent from $(D_s, s \le t)$.*

**Remark 3.5.** At first this seems insane. For instance, the server is working at rate $\mu$; yet the output is at rate $\lambda$! The explanation is that since there is an equilibrium, what comes in must equal what goes out. This makes sense from the point of view of the system, but is hard to comprehend from the point of view of the individual worker.

The independence property also doesn't look reasonable. For instance if no completed service in the last 5 hours surely the queue is empty? It turns out we have learnt nothing about the length of the queue.

**Remark 3.6.** Note that the assumption that the queue is in equilibrium is essential. If we drop it, then clearly the statement of the theorem fails to hold. Indeed, suppose instead that $X_0 = 5$. Then the first departure will happen at an exponential time of parameter $\mu$ and not $\lambda$.

**Proof.** The proof consists of a really nice time-reversal argument. Recall that $X$ is a birth and death chain and has an invariant distribution. So at equilibrium, $X$ is reversible: thus for a given $T > 0$, if $\hat{X}_t = X_{T-t}$ we know that $(\hat{X}_t, 0 \le t \le T)$ has the same distribution as $(X_t, 0 \le t \le T)$. Hence $\hat{X}$ experiences a jump of size $+1$ at constant rate $\lambda$. But note that $\hat{X}$ has a jump of size $+1$ at time $t$ if and only a customer departs the queue at time $T - t$. Therefore, departures from $X$ become arrivals for $\widehat{X}$. Since the time reversal of a Poisson process is a Poisson process, we deduce that $(D_t, t \le T)$ is itself a Poisson process with rate $\lambda$.

Thus we showed that for all $T > 0$ the process $D$ restricted to $[0, T]$ is a Poisson process of rate $\lambda$. To show that $D$ is a Poisson process on $\mathbb{R}_+$ we will use part (c) of Theorem 1.6. Indeed, for any finite collection $0 \le t_1 \le t_2 \le \ldots \le t_k$ find $T$ such that $t_k \le T$. Then since $(D_t, t \le T)$ is a Poisson process, the increments $D_{t_i} - D_{t_{i-1}}$ are independent with the Poisson distribution.

For the last assertion of the theorem, it is obvious that $X_0$ is independent from arrivals between time 0 and $T$. Reversing the direction of time this shows that $X_T$ is independent from departures between 0 and $T$. □

**Remark 3.7.** The proof remains valid for any queue at equilibrium when the queue length is a birth and death chain, e.g. for an $M/M/\infty$ queue for arbitrary values of the parameters.

**Example 3.8.** In a CD shop with many lines the service rate of the cashiers is 2 per minute. Customers spend £10 on average. How many sales do they make on average?

That really depends on the rate at which customers enter the shop, while $\mu$ is basically irrelevant so long as $\mu$ is larger than the arrival rate $\lambda$. If $\lambda = 1$ per minute, then the answer would be $60 \times 1 \times 10 = 600$.

## 3.5 Queues in tandem

Suppose that there is a first $M/M/1$ queue with parameters $\lambda$ and $\mu_1$. Upon service completion, customers immediately join a second single-server queue where the rate of service is $\mu_2$. For which values of the parameters is the chain transient or recurrent? What about equilibrium?

**Theorem 3.9.** *Let $X_t, Y_t$ denote the queue lengths in the first and second queue respectively. The process $(X, Y)$ is a positive recurrent Markov chain if and only if $\lambda < \mu_1$ and $\lambda < \mu_2$. In this case the invariant distribution is given by*

$$\pi(m, n) = (1 - \rho_1)\rho_1^m(1 - \rho_2)\rho_2^n$$

*where $\rho_1 = \lambda/\mu_1$ and $\rho_2 = \lambda/\mu_2$. In other words, $X_t$ and $Y_t$ are independent at equilibrium and are distributed according to shifted geometric random variables with parameters $1 - \rho_1, 1 - \rho_2$.*

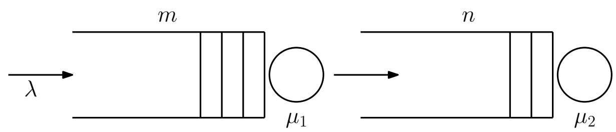



Figure 2: M/M/1 queues in series

**Proof.** We first compute the rates. From $(m, n)$ the possible transitions are

$$(m, n) \to \begin{cases} (m + 1, n) & \text{with rate } \lambda \\ (m - 1, n + 1) & \text{with rate } \mu_1 \text{ if } m \geq 1 \\ (m, n - 1) & \text{with rate } \mu_2 \text{ if } n \geq 1. \end{cases}$$

We can check by direct computation that $\pi Q = 0$ if and only if $\pi$ has the desired form. Moreover, the rates are bounded so almost surely the chain does not explode. Hence it is positive recurrent.

An alternative, more elegant or conceptual proof, uses Burke's theorem. Indeed, the first queue is an $M/M/1$ queue so no positive recurrence is possible unless $\lambda < \mu_1$. In this case we know that the equilibrium distribution is $\pi^1(m) = (1 - \rho_1)\rho_1^m$. Moreover we know by Burke's theorem that (at equilibrium) the departure process is a Poisson process with rate $\lambda$. Hence when the first queue is in equilibrium, the second queue is also an $M/M/1$ queue. Thus no equilibrium is possible unless $\lambda < \mu_2$ as well. In which case the equilibrium distribution of $Y$ is $\pi^2(n) = (1 - \rho_2)\rho_2^n$. It remains to check independence. Intuitively this is because $Y_t$ depends only on $Y_0$ and the departure process $(D_s, s \leq t)$. But this is independent of $X_t$ by Burke's theorem.

More precisely, if $X_0 \sim \pi^1$ and $Y_0 \sim \pi^2$ are independent, then Burke's theorem implies that the distribution of $(X_t, Y_t)$ is still given by two independent random variables with distributions $\pi^1$ and $\pi^2$. Hence, since $(X, Y)$ is irreducible, it follows from Remark 2.20 that the product of $\pi^1$ and $\pi^2$ is the invariant distribution of $(X, Y)$. □

**Remark 3.10.** The random variables $X_t$ and $Y_t$ are independent at equilibrium for a fixed time $t$, but the *processes* $(X_t, t \geq 0)$ and $(Y_t, t \geq 0)$ cannot be independent: indeed, $Y$ has a jump of size $+1$ exactly when $X$ has a jump of size $-1$.

**Remark 3.11.** You may wonder about transience or null recurrence. It is easy to see that if $\lambda > \mu_1$, or if $\lambda > \mu_2$ and $\lambda < \mu_1$ then the queue will be transient. The equality cases are delicate. For instance if you assume that $\lambda = \mu_1 = \mu_2$, it can be shown that $(X, Y)$ is recurrent. Basically this is because the jump chain is similar to a two-dimensional simple random walk, which is recurrent. However, with three or more queues in tandem this is no longer the case: essentially because a simple random walk in $\mathbb{Z}^d$ is transient for $d \geq 3$.

## 3.6 Jackson Networks

Suppose we have a network of $N$ single-server queues. The arrival rate into each queue is $\lambda_i, 1 \leq i \leq N$. The service rate into each queue is $\mu_i$. Upon service completion, each customer can either move to queue $j$ with probability $p_{ij}$ or exit the system with probability $p_{i0} := 1 - \sum_{j \geq 1} p_{ij}$. We assume that $p_{i0}$ is positive for all $1 \leq i \leq N$, and that $p_{ii} = 0$. We also assume that the system is irreducible in the sense that if a customer arrives in queue $i$ it is always possible for him to visit queue $j$ at some later time, for arbitrary $1 \leq i, j \leq N$.

Formally, the Jackson network is a Markov chain on $S = \mathbb{N} \times \ldots \times \mathbb{N}$ ($N$ times), where if $x = (x_1, \ldots, x_N)$ then $x_i$ denotes the number of customers in queue $i$. If $e_i$ denotes the vector with zeros everywhere except 1 in the $i$th coordinate, then

$$\begin{cases} q(n, n+e_i) & = \lambda_i \\ q(n, n+e_j-e_i) & = \mu_i p_{ij} \text{ if } n_i \geq 1 \\ q(n, n-e_i) & = \mu_i p_{i0} \text{ if } n_i \geq 1 \end{cases}$$

What can be said about equilibrium in this case? The problem seems very difficult to approach: the interaction between the queues destroys independence. Nevertheless we will see that we will get some surprisingly explicit and simple answers. The key idea is to introduce quantities, which we will denote by $\bar{\lambda}_i$, which we will later show to be the *effective rate* at which customers enter queue $i$. We can write down a system of equations that these numbers must satisfy, called the *traffic equations*, which is as follows:

**Definition 3.12.** We say that a vector $(\bar{\lambda}_1, \ldots, \bar{\lambda}_N)$ satisfies the *traffic equations* if for all $i \leq N$ we have

$$\bar{\lambda}_i = \lambda_i + \sum_{j \neq i} \bar{\lambda}_j p_{ji}. \tag{3.3}$$

The idea of (3.3) is that the effective arrival rate into queue $i$ consists of arrivals from outside the system (at rate $\lambda_i$) while arrivals from within the system, from queue $j$ say, should take place at rate $\bar{\lambda}_j p_{ji}$. The reason for this guess is related to Burke's theorem: as the effective output rate of this queue should be the same as the effective input rate.

**Lemma 3.13.** *There exists a unique solution to the traffic equations* (3.3).

**Proof.** *Existence*: Observe that the matrix $P = (p_{ij})$ defines a stochastic matrix on $\{0, \ldots, N\}$. The corresponding (discrete) Markov Chain $(Z_n)$ is transient in the sense that it is eventually absorbed at zero. Suppose $\mathbb{P}(Z_0 = i) = \lambda_i / \lambda$, for $1 \leq i \leq N$, where $\lambda = \sum_i \lambda_i$. Since $Z$ is transient, the number of visits $V_i$ to state $i$ by $Z$ satisfies $\mathbb{E}[V_i] < \infty$. But observe that

$$\begin{aligned} \mathbb{E}[V_i] &= \mathbb{P}(Z_0 = i) + \sum_{n=0}^{\infty} \mathbb{P}(Z_{n+1} = i) = \frac{\lambda_i}{\lambda} + \sum_{n=0}^{\infty} \sum_{j=1}^{N} \mathbb{P}(Z_n = j; Z_{n+1} = i) \\ &= \frac{\lambda_i}{\lambda} + \sum_{n=0}^{\infty} \sum_{j=1}^{N} \mathbb{P}(Z_n = j) p_{ji} = \frac{\lambda_i}{\lambda} + \sum_{j=1}^{N} p_{ji} \mathbb{E}[V_j]. \end{aligned}$$

Multiplying by $\lambda$, we see that if $\bar{\lambda}_i = \lambda \mathbb{E}[V_i]$, then

$$\bar{\lambda}_i = \lambda_i + \sum_{j=1}^{N} \bar{\lambda}_j p_{ji}$$

which is the same thing as (3.3) as $p_{ii} = 0$.

*Uniqueness*: see example sheet 3. □

We come to the main theorem of this section. This frequently appears in lists of the most useful mathematical results for industry.

**Theorem 3.14.** *(Jackson's theorem, 1957). Assume that the traffic equations have a solution $\bar{\lambda}_i$ such that $\bar{\lambda}_i < \mu_i$ for every $1 \le i \le N$. Then the Jackson network is positive recurrent and*

$$\pi(n) = \prod_{i=1}^{N} (1 - \bar{\rho}_i)\bar{\rho}_i^{n_i}$$

*defines an invariant distribution, where $\bar{\rho}_i = \bar{\lambda}_i/\mu_i$. At equilibrium, the processes of departures (to the outside) from each queue form independent Poisson processes with rates $\bar{\lambda}_i p_{i0}$.*

**Remark 3.15.** At equilibrium, the queue lengths $X_t^i$ are thus independent for a fixed $t$. This is extremely surprising given how much queues interact.

**Proof.** This theorem was proved relatively recently, for two reasons. One is that it took some time before somebody made the bold proposal that queues could be independent at equilibrium. The second reason is that in fact the equilibrium is non-reversible, which always makes computations vastly more complicated a priori. As we will see these start with a clever trick: we will see that there is a partial form of reversibility, in the sense of the Partial Balance Equations of the following lemma.

**Lemma 3.16.** *Suppose that $X_t$ is a Markov chain on some state space $S$, and that $\pi(x) \ge 0$ for all $x \in S$. Assume that for each $x \in S$ we can find a partition of $S \setminus \{x\}$, into say $S_1^x, \dots$ such that for all $i \ge 1$*

$$\sum_{y \in S_i^x} \pi(x)q(x,y) = \sum_{y \in S_i^x} \pi(y)q(y,x). \tag{3.4}$$

*Then $\pi$ is an invariant measure.*

**Definition 3.17.** The equations (3.4) are called the Partial Balance Equations.

**Proof.** The assumptions say that for each $x$ you can group the state space into clumps such that the flow from $x$ to each clump is equal to the flow from that clump to $x$. It is reasonable that this implies $\pi$ is an invariant measure.

The formal proof is easy: indeed,

$$\begin{aligned}\sum_x \pi(x)q(x,y) &= \sum_{x \ne y} \pi(x)q(x,y) + \pi(y)q(y,y) = \sum_i \sum_{x \in S_i^y} \pi(x)q(x,y) + \pi(y)q(y,y) \\ &= \sum_i \sum_{x \in S_i^y} \pi(y)q(y,x) + \pi(y)q(y,y) = \pi(y)\sum_x q(y,x) - \pi(y)q(y,y) + \pi(y)q(y,y) = 0,\end{aligned}$$

so $\pi Q(y) = 0$ for all $y \in S$. □

Here we apply this lemma as follows. Let $\pi(n) = \prod_{i=1}^N \bar{\rho}_i^{n_i}$ (a constant multiple of what is in the theorem). Then define

$$\widetilde{q}(n,m) = \frac{\pi(m)}{\pi(n)} q(m,n).$$

We will check that summing over an appropriate partition of the state space, $\sum_m q(n,m) = \sum_m \tilde{q}(n,m)$ which implies the partial balance equations.

Let

$$\mathcal{A} = \{e_i; 1 \le i \le N\}.$$

Thus if $n \in S$ is a state and $m \in \mathcal{A}$ then $n+m$ denotes any possible state after arrival of a customer in the system at some queue.

Let

$$\mathcal{D}_j = \{e_i - e_j; i \neq j\} \cup \{-e_j\}.$$

Thus if $n \in S$ is a state and $m \in \mathcal{D}_j$ then $n+m$ denotes any possible state after departure of a customer from queue $j$.

We will show: for all $n \in S$,

$$\sum_{m \in \mathcal{D}_j} q(n, n+m) = \sum_{m \in \mathcal{D}_j} \widetilde{q}(n, n+m) \tag{3.5}$$

$$\sum_{m \in \mathcal{A}} q(n, n+m) = \sum_{m \in \mathcal{A}} \widetilde{q}(n, n+m) \tag{3.6}$$

which implies that $\pi$ satisfies the partial balance equations and is thus invariant.

For the proof of (3.5), note that if $m \in \mathcal{D}_j$ then $q(n, n+m) = \mu_j p_{j0}$ if $m = -e_j$, and $q(n, n+m) = \mu_j p_{ji}$ if $m = e_i - e_j$. Thus the left hand side of (3.5) is

$$\sum_{m \in \mathcal{D}_j} q(n,m) = \mu_j p_{j0} + \sum_{i \neq j} \mu_j p_{ji} = \mu_j$$

which makes sense as services occur at rate $\mu_j$.

Now,

$$\widetilde{q}(n, n+e_i-e_j) = \frac{\pi(n+e_i-e_j)}{\pi(n)} q(n+e_i-e_j, n) = \frac{\bar{\rho}_i}{\bar{\rho}_j} \times \mu_i p_{ij} = \frac{\bar{\lambda}_i/\mu_i}{\bar{\rho}_j} \mu_i p_{ij} = \frac{\bar{\lambda}_i p_{ij}}{\bar{\rho}_j}.$$

Also,

$$\widetilde{q}(n, n-e_j) = \frac{\pi(n-e_j)}{\pi(n)} q(n-e_j, n) = \frac{\lambda_j}{\bar{\rho}_j}$$

We deduce that the right hand side of (3.5) is given by

$$\begin{aligned} \sum_{m \in \mathcal{D}_j} \widetilde{q}(n,m) &= \frac{\lambda_j}{\bar{\rho}_j} + \sum_{i \neq j} \frac{\bar{\lambda}_i p_{ij}}{\bar{\rho}_j} \\ &= \frac{\bar{\lambda}_j}{\bar{\rho}_j} \qquad \text{(by traffic equations)} \\ &= \mu_j, \end{aligned}$$

as desired. We now turn to (3.6). The left hand side is

$$\sum_{m \in \mathcal{A}} q(n, n+m) = \sum_i \lambda_i.$$

For the right hand side, we observe that

$$\tilde{q}(n, n+e_i) = \frac{\pi(n+e_i)}{\pi(n)} q(n+e_i, n) = \bar{\rho}_i \times \mu_i p_{i0} = \frac{\bar{\lambda}_i}{\mu_i} \times \mu_i p_{i0} = \bar{\lambda}_i p_{i0}.$$

Hence the right hand side of (3.6) is given by

$$\begin{aligned}
\sum_{m \in \mathcal{A}} \tilde{q}(n, n+m) &= \sum_i \bar{\lambda}_i p_{i0} = \sum_i \bar{\lambda}_i (1 - \sum_j p_{ij}) = \sum_i \bar{\lambda}_i - \sum_j \sum_i \bar{\lambda}_i p_{ij} \\
&= \sum_i \bar{\lambda}_i - \sum_j (\bar{\lambda}_j - \lambda_j) \qquad \text{(by traffic equations)} \\
&= \sum_j \lambda_j,
\end{aligned}$$

as desired. So $\pi$ is an invariant distribution. Since the rates are bounded, there can be no explosion and it follows that, if $\bar{\rho}_i < 1$ for every $i \geq 1$, we get an invariant distribution for the chain and hence it is positive recurrent.

For the claim concerning the departures from the queue, see Example Sheet 3. □

## 3.7 Non-Markov queues: the $M/G/1$ queue.

Consider an $M/G/1$ queue: customers arrive in a Markovian way (as a Poisson process with rate $\lambda$) to a single-server queue. The service time of the $n$th customer is a random variable $\xi_n \geq 0$, and we only assume that the $(\xi_n)_{n \geq 1}$ are i.i.d.

As usual we will be interested in the queue length $X_t$, which this time is no longer a Markov chain. What hope is there to study its long-term behaviour without a Markov assumption? Fortunately there is a hidden Markov structure underneath – in fact, we will discover two related Markov processes. Let $D_n$ denote the departure time of the $n$th customer.

**Proposition 3.18.** *The process $(X(D_n), n \geq 1)$ forms a (discrete) Markov chain with transition probabilities given by*

$$\begin{pmatrix}
p_0 & p_1 & p_2 & \dots & & \\
p_0 & p_1 & p_2 & \dots & & \\
0 & p_0 & p_1 & p_2 & \dots & \\
0 & 0 & p_0 & p_1 & p_2 & \dots
\end{pmatrix}$$

*where $p_k = \mathbb{E}\big[\exp(-\lambda\xi)(\lambda\xi)^k/k!\big]$, for all $k \geq 0$.*

**Remark 3.19.** The form of the matrix is such that the first row is unusual. The other rows are given by the vector $(p_0, p_1, \dots)$ which is pushed to the right at each row.

**Proof.** Assume $X(D_n) > 0$. Then the $(n+1)$-th customer begins his service immediately at time $D_n$. During his service time $\xi_{n+1}$, a random number $A_{n+1}$ of customers arrive in the queue. Then we have

$$X(D_{n+1}) = X(D_n) + A_{n+1} - 1.$$

If however $X(D_n) = 0$, then we have to wait until the $(n+1)$-th customer arrives. Then during his service, a random number $A_{n+1}$ of customers arrive, and we have

$$X(D_{n+1}) = X(D_n) + A_{n+1}.$$

Either way, by the Markov property of the Poisson process of arrivals, the random variables $A_n$ are i.i.d. and, given $\xi_n$, $A_n$ is Poisson $(\lambda\xi_n)$. Hence

$$\mathbb{P}(A_n = k) = \mathbb{E}[\mathbb{P}(A_n = k|\xi_n)] = \mathbb{E}\Big[\exp(-\lambda\xi)(\lambda\xi)^k/k!\Big] = p_k$$

as in the statement. The result follows. □

We write $1/\mu = \mathbb{E}[\xi]$, and call $\rho = \lambda/\mu$ the traffic intensity. We deduce the following result:

**Theorem 3.20.** *If $\rho \le 1$ then the queue is recurrent: i.e., it will empty out almost surely. If $\rho > 1$ then it is transient, meaning that there is a positive probability that it will never empty out.*

Before proving the theorem we state and prove a result about transience and recurrence of 1-dimensional random walks.

**Lemma 3.21.** *Let $(\xi_i)_{i\ge 1}$ be i.i.d. integer valued random variables and let $S_n = \xi_1 + \ldots + \xi_n$ be the corresponding random walk starting from 0. Suppose that $\mathbb{E}[|\xi_1|] < \infty$. Then $S$ is recurrent if and only if $\mathbb{E}[\xi_1] = 0$.*

**Proof.** By the strong law of large numbers if $\mathbb{E}[\xi_1] > 0$, then $S_n \to \infty$ as $n \to \infty$ almost surely. Similarly if $\mathbb{E}[\xi_1] < 0$, then $S_n \to -\infty$ as $n \to \infty$ almost surely. Hence if $\mathbb{E}[\xi_1] \ne 0$, then the walk is transient.

Suppose now that $\mathbb{E}[\xi_1] = 0$. By the strong law of large numbers, we then get that almost surely

$$\frac{S_n}{n} \to 0 \quad \text{as } n \to \infty.$$

Let $\varepsilon > 0$. Then for $n$ sufficiently large we get that

$$\mathbb{P}\left(\max_{i\le n} |S_i| \le \varepsilon n\right) \ge \frac{1}{2}. \tag{3.7}$$

Let $G_n(x)$ denote the expected number of visits to $x$ by time $n$, i.e.

$$G_n(x) = \mathbb{E}_0\left[\sum_{k=0}^{n} \mathbf{1}(X_k = x)\right].$$

Then clearly we have that for all $x$

$$G_n(x) \le G_n(0).$$

Combining this with (3.7) we get

$$2n\varepsilon G_n(0) \ge \sum_{|x|\le n\varepsilon} G_n(x) \ge \frac{1}{2}n.$$

Hence we obtain

$$G_n(0) \ge \frac{1}{4\varepsilon}.$$

Letting $\varepsilon \to 0$ shows that $G_n(0) \to \infty$ as $n \to \infty$, thus establishing recurrence. □

**Proof.** We will give two proofs because they are both instructive. The first one is to use the previous proposition. Of course $X$ is transient/recurrent in the sense of the theorem if and only if $X(D_n)$ is transient/recurrent (in the sense of Markov chains). But note that while $X(D_n) > 0$ it has the same transition probabilities as a random walk on the integers $\mathbb{Z}$ with step distribution $A_n - 1$. Note that

$$\mathbb{E}[A_n] = \sum_k \mathbb{E}[A_n \mathbf{1}(\xi_n = k)] = \sum_k \mathbb{E}[A_n \mid \xi_n = k]\, \mathbb{P}(\xi_n = k) = \sum_k \lambda k \mathbb{P}(\xi_n = k) = \lambda \mathbb{E}[\xi] = \rho.$$

Hence we can apply Lemma 3.21 to get that the walk is transient if and only if $\mathbb{E}[A_n - 1] > 0$ or equivalently $\mathbb{E}[A_n] > 1$. So the result follows.

For the second proof we will uncover a second Markov structure, which is a branching process. Call a customer $C_2$ an offspring of customer $C_1$ if $C_2$ arrives during the service of $C_1$. This defines a family tree. By the definition of the $M/G/1$ queue, the number of offsprings of each customer is i.i.d. given by $A_n$. Hence the family tree is a *branching process*. Now, the queue empties out if and only if the family tree is finite. As we know from branching process theory, this is equivalent to $\mathbb{E}[A_n] \le 1$ or $\rho \le 1$, since $\mathbb{E}[A_n] = \rho$. □

**Remark 3.22.** Note that the random walk in the above proof with step distribution $A_n - 1$ is positive recurrent if and only if $\rho < 1$. Hence, when $\rho < 1$, the Markov chain $(X(D_n))$ is positive recurrent.

**Definition 3.23.** The *busy period* is the time period measured between the time that a customer arrives to an empty system until the time a customer departs leaving behind an empty system.

As an application of the branching process argument used in the proof of Theorem 3.20 we give the following example:

**Example 3.24.** The length of the busy period $B$ of the $M/G/1$ queue with $\lambda < \mu$ satisfies

$$\mathbb{E}[B] = \frac{1}{\mu - \lambda}. \tag{3.8}$$

We start by explaining why $\mathbb{E}[B] < \infty$. If $\rho < 1$, then by Remark 3.22, the Markov chain $(X(D_n))$ is positive recurrent. Let $T$ be the length of an excursion of $(X(D_n))$. Then we get

$$\mathbb{E}[B] \le \mathbb{E}\left[\sum_{i=1}^{T} (D_i - D_{i-1})\right].$$

Since the event $\{T \le n\}$ is independent of the service times after the $n$-th customer has departed, we get that

$$\mathbb{E}[B] \le \mathbb{E}[T]\, \mathbb{E}[\xi_1] < \infty.$$

To prove (3.8), we will adopt the branching process point of view. Let $A_1$ denote the number of offspring of the root individual. Then we can write

$$B = \xi_1 + \sum_{i=1}^{A_1} B_i$$

where $B_i$ is the length of the busy period associated with the individuals forming the $i$th subtree attached to the root. Note that $A_1$ and $\xi_1$ are NOT independent. Nevertheless, given $A_1$ and $\xi_1$, the $B_j$ are independent and distributed as $B$. Thus

$$\begin{aligned}\mathbb{E}[B] &= \mathbb{E}[\xi] + \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{A_1} B_i \;\middle|\; A_1, \xi_1\right]\right] \\ &= \mathbb{E}[\xi] + \mathbb{E}[A_1 \mathbb{E}[B]] \\ &= \mathbb{E}[\xi] + \rho \mathbb{E}[B]\,.\end{aligned}$$

Hence

$$\mathbb{E}[B] = \frac{\mathbb{E}[\xi]}{1-\rho}$$

and the result follows after some simplifications.

**Remark 3.25.** The connection between trees and queues is general. Since queues can be described by random walks (as we saw) this yields a general connection between branching processes and random walks. This is a very powerful tool to describe the geometry of large random trees. Using related ideas, David Aldous constructed a "scaling limit" of large random trees, called the *Brownian continuum random tree*, in the same manner that simple random walk on the integers can be rescaled to an object called *Brownian motion*.

# 4 Renewal Theory

## 4.1 Introduction

To explain the main problem in this section, consider the following example. Suppose buses arrive every 10 minutes on average. You go to a bus stop. How long will you have to wait?

Natural answers are 5 minutes or 10 minutes. This is illustrated by the following cases: if buses arrive exactly every 10 minutes, we probably arrive at a time which is uniformly distributed in between two successive arrivals, so we expect to wait 5 minutes. But if buses arrive after exponential random variables, thus forming a Poisson process, we know that the time for the next bus after any time $t$ will be an Exponential random variable with mean 10 minutes, by the Markov property.

We see that the question is ill-posed: more information is needed, but it is counter intuitive that this quantity appears to be so sensitive to the distribution we choose. To see more precisely what is happening, we introduce the notion of a renewal process.

**Definition 4.1.** Let $(\xi_i, i \geq 1)$ be i.i.d. non-negative random variables with $\mathbb{P}(\xi > 0) > 0$. We set $T_n = \sum_{i=1}^n \xi_i$ and

$$N_t = \max\{n \geq 0 : T_n \leq t\}.$$

The process $(N_t, t \geq 0)$ is called the renewal process associated with $\xi_i$.

We think of $\xi_i$ as the interval of time separating two successive renewals; $T_n$ is the time of the $n$-th renewal and $N_t$ counts the number of renewals up to time $t$.

**Remark 4.2.** Since $\mathbb{P}(\xi > 0) > 0$ we have that $N_t < \infty$ a.s. Moreover, one can see that $N_t \to \infty$ as $t \to \infty$ a.s.

### 4.2 Elementary renewal theorem

The first result, which is quite simple, tells us how many renewals have taken place by time $t$ when $t$ is large.

**Theorem 4.3.** *If $1/\lambda = \mathbb{E}[\xi] < \infty$ then we have as $t \to \infty$*

$$\frac{N_t}{t} \to \lambda \text{ a.s.}; \quad \text{and} \quad \frac{\mathbb{E}[N_t]}{t} \to \lambda.$$

**Proof.** We only prove the first assertion here. (The second is more delicate than it looks). We note that we have the obvious inequality:

$$T_{N(t)} \leq t < T_{N(t)+1}.$$

In words $t$ is greater than the time since the last renewal before $t$, while it is smaller than the first renewal after $t$. Dividing by $N(t)$, we get

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{T_{N(t)+1}}{N(t)}.$$

We first focus on the term on the left hand side. Since $N(t) \to \infty$ a.s. and since $T_n/n \to \mathbb{E}[\xi] = 1/\lambda$ by the law of large numbers, this term converges to $1/\lambda$. The same reasoning applies to the term on the right hand side. We deduce, by comparison, that almost surely

$$\frac{t}{N(t)} \to \frac{1}{\lambda} \quad \text{as } t \to \infty$$

and the result follows. □

### 4.3 Size biased picking

Suppose $X_1, \dots, X_n$ are i.i.d. and positive. Let $S_i = X_1 + \dots X_i; 1 \leq i \leq n$. We use the points $S_i/S_n, 1 \leq i \leq n$ to tile the interval $[0,1]$. This gives us a partition of the interval $[0,1]$ into $n$ subintervals, of size $Y_i = X_i/S_n$.

Suppose $U$ is an independent uniform random variable in $(0,1)$, and let $\widehat{Y}$ denote the length of the interval containing $U$. What is the distribution of $\widehat{Y}$? A first natural guess is that all the intervals are symmetric so we might guess $\widehat{Y}$ has the same distribution as $Y = Y_1$, say. However this naive guess turns out to be wrong. The issue is that $U$ tends to fall in bigger intervals than in smaller ones. This introduces a bias, which is called a *size-biasing* effect. In fact, we will show that

$$\mathbb{P}(\widehat{Y} \in dy) = ny\mathbb{P}(Y \in dy).$$

Indeed, we have

$$\begin{aligned}\mathbb{P}\Big(\widehat{Y} \in dy\Big) &= \sum_{i=1}^n \mathbb{P}\left(\widehat{Y} \in dy, \frac{S_{i-1}}{S_n} \leq U < \frac{S_i}{S_n}\right) = \sum_{i=1}^n \mathbb{P}\left(\frac{X_i}{S_n} \in dy, \frac{S_{i-1}}{S_n} \leq U < \frac{S_i}{S_n}\right) \\ &= \sum_{i=1}^n \mathbb{E}\left[\frac{X_i}{S_n}\mathbf{1}\left(\frac{X_i}{S_n} \in dy\right)\right] = \sum_{i=1}^n y\mathbb{P}\left(\frac{X_i}{S_n} \in dy\right) = ny\mathbb{P}(Y \in dy)\,.\end{aligned}$$

Note the factor $y$ accounts for the fact that if there is an interval of size $y$ then the probability $U$ will fall in it is just $y$.

More generally we introduce the following notion.

**Definition 4.4.** Let $X$ be a nonnegative random variable with law $\mu$, and suppose $\mathbb{E}[X] = m < \infty$. Then the size-biased distribution $\widehat{\mu}$ is the probability distribution given by

$$\widehat{\mu}(dy) = \frac{y}{m}\mu(dy).$$

A random variable $\widehat{X}$ with that distribution is said to have the size-biased distribution of $X$.

**Remark 4.5.** Note that this definition makes sense because $\int_0^\infty \widehat{\mu}(dy) = \int_0^\infty (y/m)\mu(dy) = \frac{m}{m} = 1$.

**Example 4.6.** If $X$ is uniform on $[0,1]$ then $\widehat{X}$ has the distribution $2xdx$ on $(0,1)$. The factor $x$ biases towards larger values of $X$.

**Example 4.7.** If $X$ is an exponential random variable with rate $\lambda$ then the size-biased distribution satisfies

$$\begin{aligned}\mathbb{P}(\widehat{X} \in dx) &= \frac{x}{1/\lambda}\lambda e^{-\lambda x}dx \\ &= \lambda^2 x e^{-\lambda x}dx\end{aligned}$$

so $\widehat{X}$ is a Gamma$(2,\lambda)$ random variable. In particular $\widehat{X}$ has the same distribution as $X_1 + X_2$, where $X_1$ and $X_2$ are two independent exponential random variables with parameter $\lambda$.

## 4.4 Equilibrium theory of renewal processes

We will now state the main theorem of this course concerning renewal processes. This deals with the long-term behaviour of renewal processes $(N_t, t \geq 0)$ with renewal distribution $\xi$, in relation to the following set of questions: for a large time $t$, how long on average until the next renewal? How long since the last renewal? We introduce the following quantities to answer these questions.

**Definition 4.8.** Let

$$A(t) = t - T_{N(t)}$$

be the age process, i.e, the time that has elapsed since the last renewal at time $t$. Let

$$E(t) = T_{N(t)+1} - t$$

be the excess at time $t$ or residual life; i.e., the time that remains until the next renewal. Finally let

$$L(t) = A(t) + E(t) = T_{N(t)+1} - T_{N(t)}$$

be the length of the current renewal.

What is the distribution of $L(t)$ for $t$ large? A naive guess might be that this is $\xi$, but as before a size-biasing phenomenon occurs. Indeed, $t$ is more likely to fall in a big renewal interval than a small one. We hence guess that the distribution of $L(t)$, for large values of $t$, is given by $\hat{\xi}$. This is the content of the next theorem.

**Definition 4.9.** A random variable $\xi$ is called *arithmetic* if $\mathbb{P}(\xi \in k\mathbb{Z}) = 1$ for some $k$ maximal with this property. If $\xi$ is not arithmetic for any $k$, then it called non-arithmetic.

**Theorem 4.10.** *Suppose that $\xi$ is non-arithmetic. Let $\mathbb{E}[\xi] = 1/\lambda$. Then*

$$L(t) \to \widehat{\xi} \tag{4.1}$$

*in distribution as $t \to \infty$. Moreover, for all $y \geq 0$,*

$$\mathbb{P}(E(t) \leq y) \to \lambda \int_0^y \mathbb{P}(\xi > x)dx. \tag{4.2}$$

*as $t \to \infty$ and the same result holds with $A(t)$ in place of $E(t)$. In fact,*

$$(L(t), E(t)) \to (\widehat{\xi}, U\widehat{\xi}) \tag{4.3}$$

*in distribution as $t \to \infty$, where $U$ is uniform on $(0,1)$ and is independent from $\widehat{\xi}$. The same result holds with the pair $(L(t), A(t))$ instead of $(L(t), E(t))$.*

**Remark 4.11.** One way to understand the theorem is that $L(t)$ has the size-based distribution $\widehat{\xi}$ and given $L(t)$, the point $t$ falls uniformly within the renewal interval of length $L(t)$. This is the meaning of the uniform random variable in the limit (4.3).

**Remark 4.12.** Let us explain why (4.2) and (4.3) are consistent. Indeed, if $U$ is uniform and $\hat{\xi}$ has the size-biased distribution then

$$\begin{aligned}
\mathbb{P}(U\hat{\xi} \leq y) &= \int_0^1 \mathbb{P}(\hat{\xi} \leq y/u)du \\
&= \int_0^1 (\int_0^{y/u} \lambda x \mathbb{P}(\xi \in dx))du \\
&= \int_0^\infty \lambda x \mathbb{P}(\xi \in dx) \int_0^1 \mathbf{1}_{\{u \leq y/x\}}du \\
&= \int_0^\infty \lambda x \mathbb{P}(\xi \in dx)(1 \wedge y/x) \\
&= \lambda \int_0^\infty (y \wedge x)\mathbb{P}(\xi \in dx).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\lambda \int_0^y \mathbb{P}(\xi > z)dz &= \lambda \int_0^y \int_z^\infty \mathbb{P}(\xi \in dx)dz \\
&= \lambda \int_0^\infty \mathbb{P}(\xi \in dx) \int_0^\infty \mathbf{1}_{\{z<y, z<x\}}dz \\
&= \lambda \int_0^\infty \mathbb{P}(\xi \in dx)(y \wedge x)
\end{aligned}$$

so the random variable $U\hat{\xi}$ indeed has the distribution function given by (4.2).

**Example 4.13.** If $\xi \sim \text{Exp}(\lambda)$ then the renewal process is a Poisson process with rate $\lambda$. The formula

$$\lambda \int_0^y \mathbb{P}(\xi > x)dx = \lambda \int_0^y e^{-\lambda x}dx = 1 - e^{-\lambda y}$$

gives us an exponential random variable for the limit of $E(t)$. This is consistent with the Markov property: in fact, $E(t)$ is an $\text{Exp}(\lambda)$ random variable for every $t \geq 0$. Also, $\widehat{\xi} = \text{Gamma}\ (2, \lambda)$ by Example 4.7. This can be understood as the sum of the exponential random variable giving us the time until the next renewal and another independent exponential random variable corresponding to the time since the last renewal. This is highly consistent with the fact that Poisson processes are time-reversible and the notion of bi-infinite Poisson process defined in Example Sheet 2.

**Example 4.14.** If $\xi$ is uniform on $(0,1)$ then for $0 \le y \le 1$

$$\mathbb{P}(E_\infty \le y) = \lambda \int_0^y \mathbb{P}(\xi > u)du = \lambda \int_0^y (1-u)du = 2(y - y^2/2).$$

**Sketch of proof of Theorem 4.10.** We now sketch a proof of the theorem, in the case where $\xi$ is a discrete random variable taking values in $\{1, 2, \ldots\}$. We start by proving (4.2) which is slightly easier.

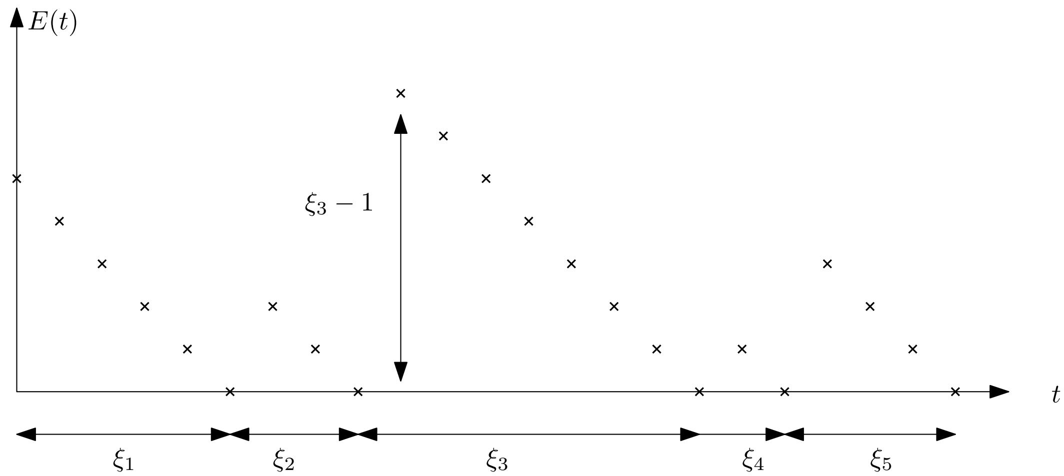



Figure 3: The residual life as a function of time in the discrete case.

Then, as suggested by Figure 3, $(E(t), t = 0, 1, \ldots)$ forms a discrete time Markov chain with transitions

$$p_{i,i-1} = 1$$

for $i \ge 1$ and

$$p_{0,n} = \mathbb{P}(\xi = n+1)$$

for all $n \ge 1$. It is clearly irreducible and recurrent, and an invariant measure satisfies:

$$\pi_n = \pi_{n+1} + \pi_0 \mathbb{P}(\xi = n+1)$$

thus by induction we deduce that

$$\pi_n := \sum_{m \ge n+1} \mathbb{P}(\xi = m)$$

is an invariant measure for this chain. This can be normalised to be a probability measure if $\mathbb{E}[\xi] < \infty$ in which case the invariant distribution is

$$\pi_n = \lambda \mathbb{P}(\xi > n).$$

We recognise the formula (4.2) in the discrete case where $t$ and $y$ are restricted to be integers. Using the assumption that $\xi$ is non-arithmetic gives that the Markov chain $E$ is aperiodic. Since it is also irreducible, we can apply the convergence to equilibrium theorem to get (4.2).

We now consider the slightly more delicate result (4.3), still in the discrete case $\xi \in \{1, 2, \ldots\}$. Of course, once this is proved, (4.1) follows. Observe that $\{(L(t), E(t)); t = 0, 1, \ldots\}$ also forms a discrete time Markov chain in the space $\mathbb{N} \times \mathbb{N}$ and more precisely in the set

$$S = \{(n, k) : 0 \le k \le n-1\}.$$

The transition probabilities are given by

$$p_{(n,k)\to(n,k-1)} = 1$$

if $k \geq 1$ and

$$p_{(n,0)\to(k,k-1)} = \mathbb{P}(\xi = k).$$

This is an irreducible recurrent chain for which an invariant measure is given by $\pi(n,k)$ where:

$$\pi(n,k-1) = \pi(n,k)$$

for $0 \leq k \leq n-1$ and

$$\pi(k,k-1) = \sum_{m=0}^{\infty} \pi(m,0)\mathbb{P}(\xi = k).$$

So taking $\pi(n,k) = \mathbb{P}(\xi = n)$ works. This can be rewritten as

$$\pi(n,k) = n\mathbb{P}(\xi = n) \times \frac{1}{n}\mathbf{1}_{\{0\leq k\leq n-1\}}.$$

After normalisation, the first factor becomes $\mathbb{P}(\hat{\xi} = n)$ and the second factor tells us that $E(t)$ is uniformly distributed on $\{0,\dots n-1\}$ given $L(t) = n$ in the limit. The theorem follows. □

## 4.5 Renewal-Reward processes

We will consider a simple modification of renewal processes where on top of the renewal structure there is a reward associated to each renewal. The reward itself could be a function of the renewal. The formal setup is as follows. Let $(\xi_i, R_i)$ denote i.i.d. pairs of random variables (note that $\xi$ and $R$ do not have to be independent) with $\xi \geq 0$ and $1/\lambda = \mathbb{E}[\xi] < \infty$. Let $N_t$ denote the renewal process associated with the $(\xi_i)$ and let

$$R_t = \sum_{i=1}^{N_t} R_i$$

denote the total reward collected up to time $t$. We begin with a result telling us about the long-term behaviour of $R_t$ which is analogous to the elementary renewal theorem.

**Proposition 4.15.** *If $\mathbb{E}[|R|] < \infty$, then as $t \to \infty$,*

$$\frac{R_t}{t} \to \lambda\mathbb{E}[R] \quad \textit{a.s. and} \quad \frac{\mathbb{E}[R_t]}{t} \to \lambda\mathbb{E}[R]\,.$$

Things are more interesting if we consider the current reward: i.e., $r(t) = \mathbb{E}\big[R_{N(t)+1}\big]$. The size-biasing phenomenon has an impact in this setup too. The equilibrium theory of renewal processes can be used to show the following fact:

**Theorem 4.16.** *As $t \to \infty$ we have*

$$r(t) \to \lambda\mathbb{E}[R\xi]\,.$$

**Remark 4.17.** The factor $\xi$ in the expectation $\mathbb{E}[R\xi]$ comes from size-biasing: the reward $R$ has been biased by the size $\xi$ of the renewal in which we can find $t$. The factor $\lambda$ is $1/\mathbb{E}[\xi]$.

## 4.6 Example: Alternating Renewal process

Suppose a machine goes on and off; on and off; etc. Each time the machine is on, it breaks down after a random variable $X_i$. Once broken it takes $Y_i$ for it to be fixed by an engineer. We assume that $X_i$ and $Y_i$ are both i.i.d. and are independent of each other. Suppose also that they have finite second moments. Let $\xi_i = X_i + Y_i$ which is the length of a full cycle. Then $\xi_i$ defines a renewal process $N_t$. What is the fraction of time the machine is on in the long-run?

We can associate to each renewal the reward $R_i$ which corresponds to the amount of time the machine was on during that particular cycle. Thus $R_i = X_i$. We deduce from Proposition 4.15 that if $R_t$ is the total amount of time that the machine was on during $(0, t)$,

$$\frac{R_t}{t} \to \frac{\mathbb{E}[X]}{\mathbb{E}[X] + \mathbb{E}[Y]} \quad \text{and} \quad \frac{\mathbb{E}[R_t]}{t} \to \frac{\mathbb{E}[X]}{\mathbb{E}[X] + \mathbb{E}[Y]}. \tag{4.4}$$

In reality there is a subtlety in deriving (4.4) from Proposition 4.15. This has to do with the fact that in the renewal reward process the reward is only collected at the end of the cycle, where as in our definition $R_t$ takes into account only the time the machine was on up to time $t$: not up to the last renewal before time $t$. The discrepancy can for instance be controlled using Theorem 4.16.

What about the probability $p(t)$ that the machine is on at time $t$? Is there a size-biasing effect taking place here as well? It can be shown no such effect needs to be considered for this question, as is suggested by (4.4) (since $\mathbb{E}[R_t] = \int_0^t p(s)ds$). One can then argue that

$$p(t) \to \frac{\mathbb{E}[X]}{\mathbb{E}[X] + \mathbb{E}[Y]} \tag{4.5}$$

as $t \to \infty$.

## 4.7 Example: busy periods of $M/G/1$ queue

Consider an $M/G/1$ queue with traffic intensity $\rho < 1$. Let $I_n$, $B_n$ denote the lengths of time during which the server is successively idle and then busy. Note that $(B_n, I_n)$ form an Alternating Renewal process. (Here it is important to consider $B_n$ followed by $I_n$ and not the other way around in order to get the renewal structure. Otherwise it is not completely obvious that the random variables are iid). It follows that if $p(t)$ is the probability that the server is idle at time $t$, then

$$p(t) \to \frac{\mathbb{E}[I]}{\mathbb{E}[B] + \mathbb{E}[I]}.$$

Now, by the Markov property of arrivals, $I_n \sim \text{Exp}(\lambda)$ so $\mathbb{E}[I] = 1/\lambda$. We have also calculated using a branching process argument (see Example 3.24)

$$\mathbb{E}[B] = \frac{1}{\mu - \lambda}.$$

We deduce that

$$p(t) \to 1 - \frac{\lambda}{\mu} \tag{4.6}$$

which is consistent with the $M/M/1$ queue in which case $p(t) \to \pi_0 = 1 - \lambda/\mu$ (letting $\pi$ denote the equilibrium distribution of the queue length).

## 4.8 Little's formula

We will now state one of applied probability's more robust theorems. First we give the definition of a regenerative process.

**Definition 4.18.** A process $(X_t : t \geq 0)$ is called *regenerative* if there exist random times $\tau_n$ such that the process regenerates, i.e. the law of the process $(X_{t+\tau_n})_{t\geq 0}$ is identical to $(X_t)_{t\geq 0}$ and independent of $(X_t)_{t\leq \tau_n}$.

**Remark 4.19.** For instance $X$ could be an M/G/1 queue and $\tau_1$ is the end of the first busy period.

For Little's formula we suppose that we are given a queue $X$ which is regenerative with regeneration times $\tau_n$. Let $N$ be the arrival process, i.e. $N_t$ counts the total number of customers that arrived before time $t$. Let $W_i$ denote the waiting time of the $i$-th customer, which is the time that he spends waiting in the queue plus the service time.

**Theorem 4.20** (Little's formula)**.** *If $\mathbb{E}[\tau_1] < \infty$ and $\mathbb{E}\left[\sum_{i=1}^{N_{\tau_1}} W_i\right] < \infty$, then almost surely the following limits exist and are deterministic:*

(a) *Long-run mean queue size*

$$L := \lim_{t\to\infty} \frac{1}{t}\int_0^t X_s\, ds.$$

(b) *Long-run average waiting time*

$$W := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i.$$

(c) *Long-run average arrival rate*

$$\lambda = \lim_{t\to\infty} \frac{N_t}{t}.$$

*Moreover, we have $L = \lambda W$.*

**Proof.** For every $n$ we define

$$Y_n = \sum_{i=1}^{N_{\tau_n}} W_i.$$

Let now $\tau_n \leq t < \tau_{n+1}$. Then we have

$$\frac{Y_n}{\tau_{n+1}} \leq \frac{1}{t}\int_0^t X_s\, ds \leq \frac{Y_{n+1}}{\tau_n}. \tag{4.7}$$

We can write $Y_{\tau_n} = \sum_{i=1}^{n}(Y_{\tau_i} - Y_{\tau_{i-1}})$, since we take $Y_0 = 0$. By the regenerative property we have that the increments $Y_{\tau_i} - Y_{\tau_{i-1}}$ are i.i.d. with $\mathbb{E}[Y_1] < \infty$ by assumption, and hence we can apply the strong law of large numbers to obtain

$$\lim_{n\to\infty} \frac{Y_n}{\tau_n} = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[\tau_1]}.$$

Therefore, taking the limits of the left and right hand side of (4.7) we get that

$$L := \lim_{t\to\infty} \frac{1}{t}\int_0^t X_s\, ds = \frac{\mathbb{E}[Y_1]}{\mathbb{E}[\tau_1]}.$$

Similarly, using the assumption that $\mathbb{E}[N_{\tau_1}] < \infty$, one can show that the following limit exists

$$\lim_{t\to\infty} \frac{N_t}{t} = \lambda.$$

For $N_{\tau_n} \le k < N_{\tau_{n+1}}$ we also have

$$\frac{Y_{\tau_n}}{N_{\tau_{n+1}}} \le \frac{1}{k}\sum_{i=1}^{k} W_i \le \frac{Y_{\tau_{n+1}}}{N_{\tau_n}}.$$

Finally we conclude that almost surely

$$\lim_{n\to\infty} \frac{Y_{\tau_n}}{N_{\tau_{n+1}}} = \frac{L}{\lambda} \quad \text{and} \quad \lim_{n\to\infty} \frac{Y_{\tau_{n+1}}}{N_{\tau_n}} = \frac{L}{\lambda}$$

and this finishes the proof. □

**Remark 4.21.** The above theorem is remarkably simple and general. Notice that there is no assumption on the inter-arrival distribution or the waiting time distribution. Moreover, there is no assumption on the number of servers. Also the customers do not have to exit the queue in the order they entered. The only assumptions are the regeneration of the queue length. From the proof above it follows that if we only assume the existence of the limits in (a) and (b) and that $\lim_{t\to\infty} X_t/t = 0$, then the formula still holds and the limit in (c) exists and is non-random. Indeed, in this case we have

$$\sum_{k=1}^{N_t - X_t} W_k \le \int_0^t X_s\, ds \le \sum_{k=1}^{N_t} W_k$$

and since $N_t/t \to \lambda$ and $X_t/t \to 0$ as $t \to \infty$, we get

$$\lim_{t\to\infty} \frac{\int_0^t X_s\, ds}{t} = \lambda W.$$

**Example 4.22.** *Waiting time in an $M/M/1$ queue.* Recall that the equilibrium distribution is $\pi_n = (1-\rho)\rho^n$ where $\rho = \lambda/\mu < 1$. Hence in that queue

$$L = \sum_n n\pi_n = \frac{1}{1-\rho} - 1 = \frac{\lambda}{\mu - \lambda}.$$

Hence by Little's theorem,

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}.$$

We recover (3.1), where in fact we had argued that the waiting time of a customer at large times $t$ was an exponential random variable with rate $\mu - \lambda$.

## 4.9 $G/G/1$ queue

Consider a $G/G/1$ queue. Let $A_n$ denote the intervals between arrival times of customers and $S_n$ their service times. Suppose that $A_n$ and $S_n$ are i.i.d. and independent of each other. It is not hard to prove the following result.

**Theorem 4.23.** *Let $1/\lambda = \mathbb{E}[A_n]$ and let $1/\mu = \mathbb{E}[S_n]$, and let $\rho = \lambda/\mu$. Then if $\rho < 1$ the queue will empty out almost surely.*

**Proof.** Let $E$ be the event that the queue never empties out. Let $A_t$ be the arrival process (number of customers arrived up to time $t$) and let $D_t$ be the departure process (number of customers served up to time $t$). Then by the elementary renewal theorem we have as $t \to \infty$

$$\frac{A_t}{t} \to \lambda \quad \text{and} \quad \frac{D_t}{t} \to \mu.$$

Hence on $E$, if $\rho > 1$, we have $D_t \gg A_t$ for $t$ large enough, which is impossible. Therefore we obtain that if $\rho < 1$, then $\mathbb{P}(E) = 0$. □

Suppose now that $\rho < 1$ for a $G/G/1$ queue and consider the times $T_n$ at which the queue is empty and a customer arrives. Then $T_n$ is a renewal sequence and by recurrence, there are infinitely many times at which the queue is empty, and hence $T_n \to \infty$ as $n \to \infty$. We can apply the strong law of large numbers to obtain that in this regime almost surely

$$\lim_{t\to\infty} \frac{X_t}{t} = 0.$$

Since $N_t$ is a renewal process with $\mathbb{E}[\xi] = 1/\lambda$, by the renewal theorem we have $N_t/t \to \lambda$ as $t \to \infty$. The existence of the limit (b) in Theorem 4.20 follows by the strong law of large numbers using the renewal times $T_n$ again.

We can now apply Remark 4.21 to get that almost surely

$$\lim_{t\to\infty} \frac{\int_0^t X_s\, ds}{t} = \lambda W.$$

# 5 Population genetics

## 5.1 Introduction

Sample the DNA of $n$ individuals from a population. What patterns of diversity do we expect to see? How much can be attributed to "random drift" vs. natural selection? In order to answer we will assume *neutral mutations* and deduce universal patterns of variation.

**Definition 5.1.** The genome is the collection of all genetic information of an individual. This information is stored on a number of chromosomes. Each consists of (usually many) genes. A gene is a piece of genetic material coding for one specific protein. Genes themselves are made up of acid bases: e.g. ATCTTAG... An *allele* is one of a number of alternative forms of the same gene. The location of a base is called *site*.

For instance, to simplify greatly, if there was a gene coding for the colour of the eye we could have the blue allele, the brown allele, etc.

To simplify, we will make a convenient abuse of language and speak of an individual when we have in mind a given gene or chromosome. In particular, for diploid populations, every member of the population has two copies of the same chromosome, which means we have two corresponding "individuals". In other words, we treat the two chromosomes in a given member of the population as two distinct individuals. So gene and individual will often mean the same thing.

## 5.2 Moran model

Our basic model of a population dynamics will be the Moran model. This is a very crude model for the evolution of a population but nevertheless captures the right essential features, and allows us to give a rigorous treatment at this level.

**Definition 5.2.** *Let $N \geq 1$. In the Moran model the population size is constant equal to $N$. At rate 1, every individual dies. Simultaneously, an individual chosen uniformly at random from the population gives birth.*

Note that the population size stays constant throughout this mechanism. In particular, we allow an individual to give birth just at the time he dies.

The Moran model can be conveniently constructed in terms of Poisson processes. In the definition of the Moran model, one can imagine that when an individual $j$ dies and individual $i$ gives birth, we can think that the offspring of $i$ is replacing $j$. By properties of Poisson processes, if the rate at which an offspring of $i$ replaces $j$ is chosen to be $1/N$ this gives us a construction of the Moran model: indeed the total rate at which $j$ dies will be $N \times (1/N) = 1$, and when this happens the individual $i$ whose offspring is replacing $j$ is chosen uniformly at random.

Thus a construction of the Moran model is obtained by considering independent Poisson processes $(N_t^{i,j}, t \geq 0)$ for $1 \leq i, j \leq N$ with rates $1/N$. When $N^{i,j}$ has a jump this means that individual $j$ dies and is replaced by an offspring of individual $i$.

**Corollary 5.3.** *The Moran model dynamics can be extended to $t \in \mathbb{R}$, by using bi-infinite Poisson processes $(N_t^{i,j}, t \in \mathbb{R})$.*

## 5.3 Fixation

Suppose at time $t = 0$, a number of $X_0 = i$ of individuals carry a different allele, called $a$, while all other $N - i$ individuals carry the allele $A$. Let $X_t = \#$ individuals carrying allele $a$ at time $t \geq 0$, using the Moran model dynamics. Let $\tau = \inf\{t \geq 0 : X_t = 0 \text{ or } N\}$. We say that $a$ fixates if $\tau < \infty$ and $X_\tau = N$. We say that there is no fixation if $\tau < \infty$ and $X_\tau = 0$.

**Theorem 5.4.** *We have that $\tau < \infty$ a.s. so these are the only two alternatives, and $\mathbb{P}(X_\tau = N | X_0 = i) = i/N$. Moreover,*

$$\mathbb{E}[\tau \mid X_0 = i] = \sum_{j=1}^{i-1} \frac{N-i}{N-j} + \sum_{j=i}^{N-1} \frac{i}{j}.$$

**Remark 5.5.** *If $p = i/N \in (0,1)$ and $N \to \infty$ then it follows that $\mathbb{E}[\tau]$ is proportional to $N$ and more precisely, $\mathbb{E}[\tau] \sim N(-p \log p - (1-p)\log(1-p))$.*

**Proof.** We begin by observing that $X_t$ is a Markov chain with

$$q_{i,i+1} = (N-i) \times i/N \tag{5.1}$$

(the first factor corresponds to an individual from the $A$ population dying, the second to choosing an individual from the $a$ population to replace him). Likewise,

$$q_{i,i-1} = i \times (N-i)/N. \tag{5.2}$$

Hence the $q_{i,i-1} = q_{i,i+1}$, $q_i = 2i(N-i)/N$, and $X_t$ is a Birth and Death chain whose jump is Simple Random Walk on $\mathbb{Z}$, absorbed at 0 and $N$. We thus get the first result from the fact that for Simple Random Walk,

$$\mathbb{P}_i(T_N < T_0) = i/N. \tag{5.3}$$

Now let us write $\tau = \sum_{j=1}^N \tau_j$ where $\tau_j$ is the total time spent at $j$. Note that

$$\mathbb{E}_j[\tau_j] = \frac{1}{q_j}\mathbb{E}_j[\# \text{ visits to } j] = \frac{1}{q_j}\mathbb{P}_j(\text{no return to } j)^{-1}$$

since the number of visits to $j$ is a geometric random variable. Now, by decoposing on the first step, the probability to not return to $j$ is given by

$$\mathbb{P}_j(\text{ no return to } j) = \frac{1}{2}\frac{1}{j} + \frac{1}{2}\frac{1}{N-j}$$

by using (5.3) on the interval $[0,j]$ and $[j,N]$ respectively. Hence

$$\mathbb{E}_j[\tau_j] = \frac{1}{q_j}\frac{2j(N-j)}{N} = 1.$$

Consequenty,

$$\begin{aligned}
\mathbb{E}_i[\tau] &= \sum_j \mathbb{E}_i[\tau_j] \\
&= \sum_i \mathbb{P}_i(X_t = j \text{ for some } t \geq 0) \times \mathbb{E}_j[\tau_j] \\
&= \sum_{j \geq i} \frac{i}{j} + \sum_{j<i} \frac{N-i}{N-j}.
\end{aligned}$$

The result follows. □

### 5.4 The infinite sites model of mutations

Consider the case of point mutations. These are mutations which change one base into another, say $A$ into $G$. When we consider a long sequence of DNA it is extremely unlikely that two mutations will affect the same base or *site*. We will make one simplifying assumption that there are infinitely many sites: i.e., no two mutations affect the same site.

Concretely, we consider the (bi-infinite) Moran model. We assume that independently of the population dynamics, every individual is subject to a mutation at rate $u > 0$, independently for all individuals (*neutral mutations*). Since we suppose that there infinitely many sites, we can safely assume that no two mutations affect the same site. It thus makes sense to ask the following question: *Sample 2 individuals from the population at time* $t = 0$. *What is the probability they carry the same allele?* More generally we can sample $n$ individuals from the population and ask how many alleles are there which are present in only one individual of the sample? Or two individuals?

The infinite sites model tells us that if we look base by base in the DNA sequence of a sample of individuals, either all bases agree in the sample, or there are two variants (but no more), since we assumed that no two mutations affect the same site.

**Definition 5.6.** Suppose we are given a table with the DNA sequences of all $n$ individuals in the sample. Let $M_j(n)$ be the number of sites where exactly $j$ individuals carry a base which differs from everyone else in the sample.

**Remark 5.7.** Note that in the above definition the quantity $M_j(n)$ counts the number of sites that have been mutated and exactly $j$ individuals carry the same mutation.

**Example 5.8.** Suppose the DNA sequences in a sample are as follows

$$\begin{array}{lcccccccccccc}
1: & \dots & A & T & T & T & C & G & G & G & T & C & \dots \\
2: & \dots & - & A & - & G & - & - & - & - & - & - & \dots \\
3: & \dots & - & - & - & - & - & - & - & - & C & - & \dots \\
4: & \dots & - & - & - & G & - & - & - & - & C & - & \dots \\
5: & \dots & - & A & - & - & - & - & - & - & - & - & \dots \\
6: & \dots & - & - & - & - & - & - & - & - & - & - & \dots \\
7: & \dots & - & - & - & - & - & - & - & - & C & - & \dots
\end{array}$$

In this example $n = 7$. To aid visualisation we have put a dash if the base is identical to that of the first individual in the sample. Hence we have $M_2(n) = 2$ (second and fourth sites) and $M_3(n) = 1$ (second to last).

Our first result tells us what happens in the (unrealistic) case where $n = N$.

**Theorem 5.9.** *Let $\theta = uN$. Then*

$$\mathbb{E}[M_j(N)] = \frac{\theta}{j}.$$

**Proof.** Mutations occur at a total rate of $u \times N = \theta$ in the time interval $(-\infty, 0]$. Suppose a mutation arises at time $-t$ $(t > 0)$ on some site. What is the chance that it affects exactly $j$ individuals in the population at time 0? Let $X_s$ denote the number of individuals carrying this mutation at time $-t + s$. Then since mutations don't affect each other, $X$ evolves like the Markov chain in the previous theorem, i.e., has the $Q$-matrix given by (5.1) and (5.2). Hence the chance that this mutation affects exactly $j$ individuals in the population at time zero is precisely $p_t(1, j)$ where $p_t(x, y)$ is the semi-group associated with the $Q$-matrix. Thus

$$\begin{aligned}
\mathbb{E}[M_j(N)] &= \int_0^\infty uN dt p_t(1, j) \\
&= \theta \mathbb{E}_1[\tau_j] \\
&= \theta \mathbb{P}_1(X_t = j \text{ for some } t \geq 0) \, \mathbb{E}_j[\tau_j] \\
&= \theta \times (1/j) \times 1,
\end{aligned}$$

as desired. □

## 5.5 Kingman's $n$-coalescent

Consider a Moran model defined on $\mathbb{R}$. Sample $n$ individuals at time 0. What is the genealogical tree of this sample? Since an individual is just a chromosome, there is just one parent for any given individual. (This is one of the advantages of making this change of perspective). Thus for any $t > 0$ there is a unique ancestor for this individual at time $-t$. As $-t$ goes further and further back in time, it may happen that the ancestor for two individuals in the population becomes the same. We speak of a *coalescence event*.

To put this on a mathematical footing, we introduce the notion of *ancestral partition*. This is a partition $\Pi_t$ of the sample (identified with $\{1, \dots, n\}$) such that $i$ and $j$ are in the same block of

$\Pi_t$ if and only if $i$ and $j$ have the same ancestor at time $-t$. One way to think about $\Pi_t$ is that there is a block for each distinct ancestor of the population at time $-t$.

How does $\Pi_t$ evolve as $t$ increases? It turns out that $\Pi_t$ forms a Markov process with values in $\mathcal{P}_n = \{\text{partitions of } \{1,\ldots,n\}\}$.

**Theorem 5.10.** $(\Pi_{Nt/2}, t \geq 0)$ *is a Markov chain in* $\mathcal{P}_n$ *with*

$$q_{\pi,\pi'} = \begin{cases} 1 & \text{if } \pi' \text{ can be obtained from } \pi \text{ by coagulating two of its blocks} \\ -\binom{k}{2} & \text{if } \pi' = \pi \text{ and } \pi \text{ has } k \text{ blocks} \\ 0 & \text{else .} \end{cases}$$

*This is Kingman's* $n$*-coalescent.*

**Proof.** It suffices to show that $\Pi_t$ is a Markov chain with rates $(2/N)q_{\pi,\pi'}$. Now, recall that each block of $\Pi_t$ is associated to an ancestor of the sample at time $-t$. The rate at which this pair of blocks coalesces is $2/N$, since if the ancestors are $i$ and $j$ at this time then the rate is equal to the sum of the rates for $N^{i,j}$ and $N^{j,i}$ in the Poisson construction of the Moran model, i.e., $2/N$. All other transitions do not take place, hence the result. □

**Properties.** We list some immediate properties of Kingman's $n$-coalescent.

1. $\Pi_0 = \{\{1\},\ldots,\{n\}\}$.
2. For $t$ sufficiently large $\Pi_t = \{\{1,\ldots,n\}\}$. The first such time is the time to the MRCA (the most recent common ancestor) of the sample.
3. $\Pi_t$ is a coalescing process. The only possible transitions involve merging a pair of blocks. Each possible pair of blocks merges at rate 1 in Kingman's $n$-coalescent (and at rate $2/N$ for $\Pi$ itself).
4. If $K_t$ is the size of Kingman's $n$-coalescent, i.e. the number of blocks at time $t$, then $K_t$ is a pure death process with rates $k \to k-1$ given by $\binom{k}{2}$. Moreover the jump chain is independent from $K_t$.

Recall Theorem 5.9 and the quantity $M_j(n)$, which is the number of sites where exactly $j$ individuals carry a base that has been mutated and is different from everyone else in the sample. Then $M_j(n)$ depends only on the mutations which intersect with the genealogical tree. In other words, we have a genealogical tree with each pair of branches coalescing at rate $2/N$. Mutations fall on the tree at rate $u > 0$ per branch and per unit length. For such a tree we proved in Theorem 5.9 that $\mathbb{E}[M_j(N)] = \theta/j$, where $\theta = uN$.

If we speed up time by $N/2$, then the genealogy becomes Kingman's $n$-coalescent and the rate of mutations becomes $\alpha = uN/2$. For any such tree Theorem 5.9 applies to give that

$$\mathbb{E}[M_j(n)] = \frac{2\alpha}{j} = \frac{\theta}{j}.$$

Therefore, we proved:

**Theorem 5.11.** *For any* $1 \leq n \leq N$*, for* $\theta = uN$*,* $\mathbb{E}[M_j(n)] = \theta/j$.

The function $j \mapsto \theta/j$ is called the site frequency spectrum of the infinite sites model.

**Example 5.12.** Biologists often measure the so-called SNP count $S_n$, or Single Nucleotide Polymorphism. This is the number of sites in the sequence for which there is some variation in the sequence, i.e. $S_n = \sum_{j=1}^n M_j(n)$. Then we deduce from the theorem above that as $n \to \infty$

$$\mathbb{E}[S_n] = \theta \left(1 + \ldots + \frac{1}{n}\right) \sim \theta \log n.$$

## 5.6 Consistency and Kingman's infinite coalescent

A further interesting property is the compatibility or *sampling consistency.* Intuitively, this means that if we have a sample of size $n$, then a subsample of size $n-1$ behaves as if we had directly $n-1$ individuals from the population. Mathematically, this can be expressed as follows. If $\pi$ is a partition of $[n] = \{1, \ldots, n\}$ then we can speak of $\pi|_{[n-1]}$, the induced partition of $[n-1]$ obtained by restricting $\pi$ to $[n-1]$.

**Proposition 5.13.** *Let $\Pi^n$ be Kingman's $n$-coalescent. Then the restriction to $[n-1]$, i.e. the process $\Pi^n|_{[n-1]}$ has the law of Kingman's $(n-1)$-coalescent.*

**Proof.** This follows directly from the construction of Kingman's $n$-coalescent by sampling from the Moran model. Alternatively it can be shown directly using the transition rates via some rather tedious calculations. □

Using the sampling consistency and Kolmogorov's extension theorem we can deduce the existence of a unique process $(\Pi_t, t \geq 0)$ taking values in partitions $\mathcal{P}$ of $\mathbb{N} = \{1, 2, \ldots\}$ such that for every $n \geq 1$, the process $\Pi|_{[n]}$ has the law of Kingman's $n$-coalescent.

**Definition 5.14.** $(\Pi_t, t \geq 0)$ *is called Kingman's infinite coalescent.*

Initially we have $\Pi_0$ consisting of infinitely many singletons. How does it look like for positive times? For instance, will it ever completely coalesce? One remarkable phenomenon with Kingman's coalescent is the following fact.

**Theorem 5.15.** *Kingman's coalescent comes down from infinity: that is, with probability one, the number of blocks of $\Pi_t$ is finite at any time $t > 0$. In particular, there is a finite time $\zeta > 0$ such that $\Pi_t = \{1, 2, \ldots\}$ for $t \geq \zeta$.*

This should be viewed as some kind of big bang event, reducing the number of blocks from infinity to finitely many in an infinitesimal amount of time.

**Proof.** Write $|\Pi_t|$ for the number of blocks. Since the events $\{|\Pi_t^n| \geq M\}$ are increasing in $n$, we get

$$\mathbb{P}(|\Pi_t| \geq M) = \lim_{n\to\infty} \mathbb{P}(|\Pi_t^n| \geq M) = \lim_{n\to\infty} \mathbb{P}\left(\sum_{j=M+1}^{n} \tau_j \geq t\right)$$

where $\tau_j$ is the time for $\Pi^n$ to drop from $j$ blocks to $j-1$ blocks. Hence $\tau_j$ is Exponential with rate $\binom{j}{2}$. By Markov's inequality

$$\mathbb{P}(|\Pi_t| \geq M) \leq \frac{1}{t}\mathbb{E}\left[\sum_{j=M+1}^{\infty} \tau_j\right] \leq \frac{1}{t} \sum_{j=M+1}^{\infty} \frac{1}{\binom{j}{2}} = \frac{2}{t} \sum_{j=M+1}^{\infty} \frac{1}{j(j-1)}$$

This tends to 0 as $M \to \infty$, so the result follows. □

## 5.7 Intermezzo: Pólya's urn and Hoppe's urn*

The proofs in this section are not examinable but we will use the conclusions later on. Consider the following urn model, due to Pólya. Initially an urn contains one white and one black ball. At each subsequent step, a ball is drawn from the urn. The ball is then put back in the urn along with a ball of the same colour. Let $X_n$ denote the number of black balls in the urn when there are $n$ balls in total in the urn. What is the limiting behaviour of $X_n$?

The first time they see this question, many people believe that $X_n/n$ will converge to 1/2 when $n \to \infty$. But the result is quite different.

**Theorem 5.16.** *We have that as $n \to \infty$, $X_n/n \to U$ almost surely, where $U$ is a uniform random variable on $(0, 1)$.*

This theorem can be thought of as a 'rich get richer' phenomenon. Initially there is a lot of randomness. There are a great variety of events that might happen during the first $n = 100$ draws say. However, once there is a large number of balls in the urn, a law of large number kicks in. For instance if the fraction is $p$ at that time, then the probability to pick a black ball will be $p$ and the probability to pick a while ball will be $1 - p$. Hence by the law of large numbers the fraction of black balls will tend to remain close to $p$ for a very long time. This reinforces itself and explain why the convergence is almost sure. We will sketch a different (rigorous) proof below.

**Proof.** We start by doing a few simple computations. What is the probability to get first $m$ black balls and then $n - m$ white balls (in that order)? We see that it is

$$\frac{1}{2}\frac{2}{3}\cdots\frac{m}{m+1} \times \frac{1}{m+2}\cdots\frac{n-m}{n+1} = \frac{m!(n-m)!}{(n+1)!}. \tag{5.4}$$

The first factor account for drawing all the black balls (whose number increase from 1 to $m - 1$ at the last draw) and the second accounts for then drawing all white balls, whose numbers increase from 1 to $n - m$ at the last draw.

The key observation is *exchangeability*: if we were to compute the probability of any other sequence of draws, also resulting in $m + 1$ black balls and $n + 1$ white balls in the urn, the probability would be unchanged. This is because the bottom of the fraction gives the number of balls in the urn (which can only increase by one at each draw) and the fraction gives the number of black or white balls currently in the urn. But this has to go increase from 1 to $m$ and from 1 to $n$ respectively, albeit at different times than in the above order. Still the product is unchanged. Hence

$$\begin{aligned}\mathbb{P}(X_{n+2} = m) &= \binom{n}{m}\frac{m!(n-m)!}{(n+1)!} \\ &= \frac{n!}{m!(n-m)!}\frac{m!(n-m)!}{(n+1)!} \\ &= \frac{1}{n+1},\end{aligned}$$

so $X_{n+2}$ is uniformly distributed over $\{1, \dots, n+1\}$. It is hence no surprise that the limit of $X_n/n$, if it exists, is uniform over $[0, 1]$.

To see why in fact the limit does exist, we assert that $X_n$ has the same dynamics as the following (seemingly very different) process: first pick a number $U \in (0, 1)$ uniformly at random. Then insert in the urn a black ball with probability $U$, and a while ball with probability $1-U$. Indeed, the probability to get $m$ black balls followed by $n-m$ white balls in that process is given by

$$\int_0^1 u^m(1-u)^{n-m}du.$$

This integral can easily be evaluated and shown to be identical to $m!(n-m)!/(n+1)!$. Clearly the probability is also invariant under permutations of the sequence, so these two processes must be identical! By the law of large numbers, it now follows that $X_n/n \to U$, almost surely. $\square$

It seems mad that the two processes considered in the proof can in fact be identical. In the Pólya urn case, there is a complex dependency phenomenon dominated by 'rich get richer'. In the second there is no such dependency – it is perhaps the most basic process of probability: i.i.d. draws, except that the parameter for drawing is itself random and is identical for all draws.

Hoppe's urn is a generalisation of Pólya's urn. This is an urn with balls of different colours of mass 1 and a single black ball of mass $\theta$. At each step, we draw from the urn (with probability proportional to the mass of the ball). If it is a coloured ball, we put back the ball in the urn along with a ball of the same colour. If it is a black ball we put it back in the urn along with a ball of a new colour.

## 5.8 Infinite Alleles Model

Consider a Moran model with $N$ individuals, defined for $t \in \mathbb{R}$. Assume that each individual is subject to a mutation at rate $u > 0$. When a mutation occurs, it is unlikely that the allelic type remains the same, or is identical to something which ever arose prior to that. Simplifying, this leads to the *Infinite Alleles Model*: we assume that each time a mutation occurs, the allelic type of the corresponding individual changes to something entirely new. (For instance, thinking of eye colour, if the type was blue before the mutation, it could change to any different colour after, say green).

The Infinite Sites Model and the Infinite Alleles Model look quite similar on the surface. However, in the Infinite Alleles Model we only look at the individual's current allelic type, and have no way of knowing or guessing the allelic type of the individual's ancestors. On the contrary this information remains accessible in the case of the Infinite Sites Model as we are given the full DNA sequence. So the main difference between the models is that we don't know if two allelic types are close or completely unrelated. We just know they are different. This is particularly appropriate in some cases where sequencing is not convenient or too expensive.

In the Infinite Alleles Model the variation in the sample is encoded by a partition $\Pi_n$ of the sample (identified, as usual, with $\{1,\dots,n\}$) such that $i$ is in the same block of $\Pi_n$ as $j$ if and only if $i$ and $j$ have the same allelic type.

**Definition 5.17.** $\Pi_n$ is called the allelic partition.

As in the case of the ISM, we introduce the quantities $A_j = A_j(n) = \#$ of distinct alleles which are carried by exactly $j$ individuals. In terms of partitions, the quantity $A_j$ counts how many sets of the partition have exactly $j$ elements.

**Example 5.18.** Suppose $n = 8$ and the eye colour of the sample is

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| blue | red | brown | green | brown | yellow | red | brown |

Then $\Pi_n$ has 4 blocks (corresponding to the five colours in the sample): $\{1\}, \{2, 7\}, \{3, 5, 8\}, \{4\}, \{6\}$. Hence $a_1 = 3, a_2 = 1, a_3 = 1$.

## 5.9 Ewens sampling formula

It turns out that we can describe the distribution of $\Pi_n$ or, equivalently, of $(A_1, \dots, A_n)$, explicitly. This is encoded in a beautiful and important formula which bears the name of Warren Ewens, who discovered it in 1972. This is also widely used by geneticists in practice.

**Theorem 5.19** (Ewens sampling formula)**.** *Let $a_j$ be such that $\sum_{j=1}^n ja_j = n$. Then*

$$\mathbb{P}(A_1 = a_1, \dots, A_n = a_n) = \frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{j=1}^n \frac{(\theta/j)^{a_j}}{a_j!}. \tag{5.5}$$

**Remark 5.20.** It is far from obvious that the right hand side adds up to one! Another way of rewriting (5.5) is

$$\mathbb{P}(A_1 = a_1, \dots, A_n = a_n) = c(\theta, n) \prod_{j=1}^n e^{-\theta/j} \frac{(\theta/j)^{a_j}}{a_j!}, \tag{5.6}$$

where $c(\theta, n)$ is a constant only depending on $\theta$ and $n$. The product appearing above is the probability that $n$ independent random variables $Z_1, \dots, Z_n$ with $Z_j \sim$Poisson$(\theta/j)$ are equal to $a_1, \dots, a_n$ respectively. Hence the probability appearing in (5.6) is $\mathbb{P}\Big(Z_1 = a_1, \dots, Z_n = a_n \,\Big|\, \sum_j jZ_j = n\Big)$.

When $n$ is large and $j$ is finite, this conditioning becomes irrelevant, and hence the distribution of $A_j$ is close to Poisson with mean $\theta/j$. So for large $n$ we have $\mathbb{P}(A_j = 0) \approx e^{-\theta/j}$.

**Remark 5.21.** An equivalent way of stating the formula is that for $\pi \in \mathcal{P}_n$ we have

$$\mathbb{P}(\Pi_n = \pi) = \frac{\theta^k}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{i=1}^k (n_i - 1)! \tag{5.7}$$

where $k$ is the number of blocks of $\pi$ and $n_i$ is the size of the $i$-th block. This is actually the version we will use below.

We now show that the two formulas (5.5) and (5.7) are equivalent. It is obvious that the distribution of $\Pi_n$ is invariant under permutation of the labels. So if $(a_j)$ is fixed such that $\sum_j ja_j = n$ and $\pi$ is a given partition having $(a_j)$ as its allele count, we have:

$$\begin{aligned}
\mathbb{P}(A_1 = a_1, \dots, A_n = a_n) &= \mathbb{P}(\Pi_n = \pi) \times \#\{\text{partitions with this allele count}\} \\
&= \frac{\theta^k}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{j=1}^k (n_j - 1)! \times n! \frac{1}{\prod_{j=1}^k n_j!} \frac{1}{\prod_{i=1}^n a_i!} \\
&= \frac{\theta^k n!}{\theta(\theta+1)\dots(\theta+n-1)} \times \frac{1}{\prod_{j=1}^k n_j} \frac{1}{\prod_{i=1}^n a_i!} \\
&= \frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{j=1}^n \frac{(\theta/j)^{a_j}}{a_j!},
\end{aligned}$$

as desired.

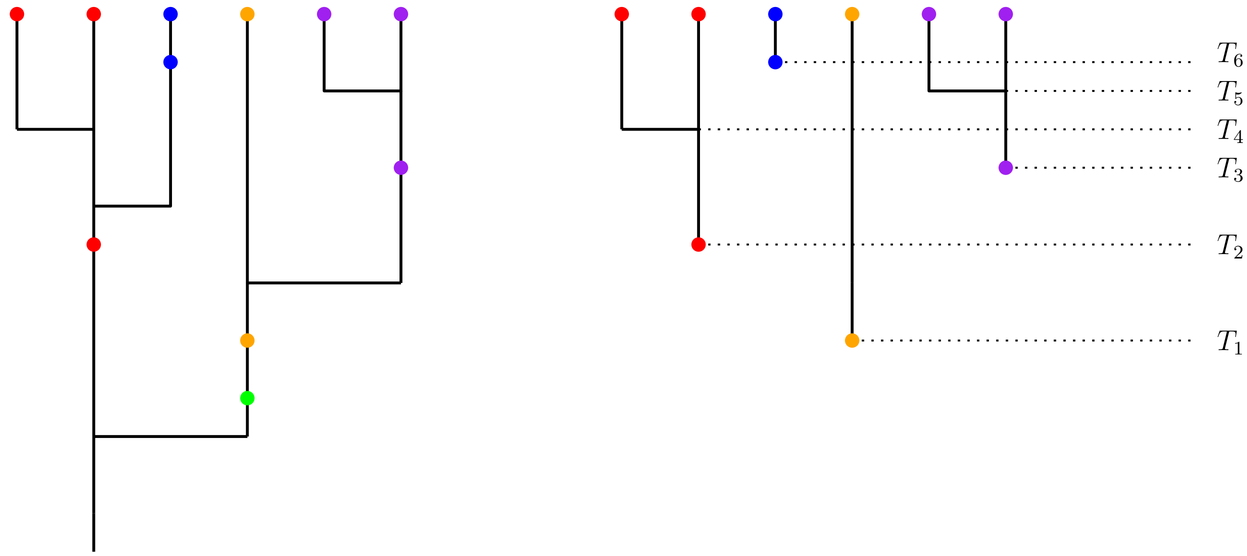



Figure 4: On the left, a Kingman $n$-coalescent, with mutations falling at rate $\theta/2$ on each lineage. An individual from the sample is coloured according to the allelic type which it carries. On the right, the same process where ancestral lineages are killed off when there is a mutation, and the definition of the times $T_1, \ldots, T_n$ in the proof.

**Proof of Theorem 5.19.** As we have done before, $(A_1, \ldots, A_n)$ depends only on the mutations which intersect the genealogical tree of the sample. Hence we may and will assume that the genealogical tree is given by Kingman's $n$-coalescent and that mutations fall on the tree at rate $\theta/2$ per unit length on each branch.

*Step 1.* We think of each mutation as a killing. Hence as the time of the coalescent evolves, branches disappear progressively, either due to coalescence or to killing caused by a mutation. Let $T_n, \ldots, T_1$ denote the successive times at which the number of branches drops from $n$ to $n-1$, then from $n-1$ to $n-2$, and so on. The key idea of the proof is to try to describe what happens in the reverse order, going from $T_1$ to $T_2$ all the way up to $T_n$.

We now consider the partition $\Pi_m$, which is the partition defined by the $m$ lineages uncovered by time $T_m$.

Between times $T_m$ and $T_{m+1}$ there are $m$ branches. At time $T_{m+1}$ we add an extra branch to the tree. This can be attached to an existing allelic group of branches (corresponding, in the time direction of the coalescent, to a coalescence event) or create a new allelic group (corresponding to a mutation event). We will now calculate the probabilities of these various possibilities, using "competition of exponential random variables": between $T_m$ and $T_{m+1}$ there were $m$ branches and at time $T_{m+1}$ we add a new branch. So there are in total $m+1$ exponential clocks with rate $\theta/2$ corresponding to mutation, and $m(m+1)/2$ clocks with rate 1, corresponding to coalescence.

Hence the probability to form a new allelic group for the new branch at time $T_m$ is

$$\mathbb{P}(\text{new block}) = \frac{(m+1)\theta/2}{(m+1)\theta/2 + m(m+1)/2} = \frac{\theta}{\theta+m}. \tag{5.8}$$

The probability to join an existing group of size $n_i$ is

$$\mathbb{P}(\text{join a group of size } n_i) = \frac{m(m+1)/2}{(m+1)\theta/2 + m(m+1)/2} \times \frac{n_i}{m} = \frac{n_i}{m+\theta}. \tag{5.9}$$

The extra factor $n_i$ comes from the fact that, given that this extra branch disappeared by coalescence, it joined a uniformly chosen existing branch, and hence joins a group of size $n_i$ with probability $n_i/m$.

*Step 2.* We observe that the above process behaves exactly as Hoppe's urn from the previous section. The "new block" or mutation event is the same as drawing the black ball of mass $\theta > 0$, while the event joining a group of size $n_i$ is identical to drawing a ball from that colour. We will now prove

$$\mathbb{P}(\Pi_n = \pi) = \frac{\theta^k}{\theta(\theta+1)\dots(\theta+n-1)}\prod_{i=1}^{k}(n_i-1)!$$

by induction on $n$, where $\pi \in \mathcal{P}_n$ is arbitrary and $k$ is the number of blocks of $\pi$. The case $n = 1$ is trivial. Now let $n \geq 2$, and let $\pi' = \pi|_{[n-1]}$. There are two cases to consider: either (a) $n$ is a singleton in $\pi$, or (b) $n$ is in a block of size $n_j$ in $\pi$. In case (a), $\pi'$ has $k-1$ blocks. Hence

$$\begin{aligned}
\mathbb{P}(\Pi_n = \pi) &= \mathbb{P}(\Pi_{n-1} = \pi') \times \frac{\theta}{\theta+n-1} \\
&= \frac{\theta^{k-1}}{\theta(\theta+1)\dots(\theta+n-2)}\prod_{i=1}^{k-1}(n_i-1)! \times \frac{\theta}{\theta+n-1} \\
&= \frac{\theta^{k}}{\theta(\theta+1)\dots(\theta+n-1)}\prod_{i=1}^{k}(n_i-1)!
\end{aligned}$$

as desired.

In case (b),

$$\begin{aligned}
\mathbb{P}(\Pi_n = \pi) &= \mathbb{P}(\Pi_{n-1} = \pi') \times \frac{n_j-1}{\theta+n-1} \\
&= \frac{\theta^{k}}{\theta(\theta+1)\dots(\theta+n-2)}\prod_{i=1;i\neq j}^{k}(n_i-1)! \times (n_j-2)! \times \frac{n_j-1}{\theta+n-1} \\
&= \frac{\theta^{k}}{\theta(\theta+1)\dots(\theta+n-1)}\prod_{i=1}^{k}(n_i-1)!
\end{aligned}$$

as desired. Either way, the formula is proved. □

**Corollary 5.22.** *Let $K_n = \#$ distinct alleles in a sample of size $n$. Then*

$$\mathbb{E}[K_n] = \sum_{i=1}^{n}\frac{\theta}{\theta+i-1} \sim \theta\log n;$$

$$\operatorname{var}(K_n) \sim \theta \log n$$

*and*

$$\frac{K_n - \mathbb{E}[K_n]}{\sqrt{\operatorname{var}(K_n)}} \to \mathcal{N}(0,1)$$

*a standard normal random variable.*

**Proof.** This follows from the Hoppe's urn representation in the proof. At step $i$ a new block is added with probability $p_i = \theta/(\theta+i-1)$. Hence $K_n = \sum_{i=1}^{n} B_i$ where $B_i$ are independent Bernoulli random variables with parameter $p_i$. The expressions for $\mathbb{E}[K_n]$, $\operatorname{var}(K_n)$ follow, and the central limit theorem comes from computing the characteristic function. □

The Central Limit Theorem is what is needed for hypothesis testing. $K_n/\log n$ is an estimator of $\theta$ which is asymptotically normal. But its standard deviation is of order $1/\sqrt{\log n}$. Eg if you want $\sigma = 10\%$ you need $n = e^{100}$, which is totally impractical...! Unfortunately, $K_n$ is a sufficient statistics for $\theta$ (see example sheet): the law of the allelic partition $\Pi_n$, given $K_n$, does not depend on $\theta$. Hence there is no information about $\theta$ beyond $K_n$, so this really is the best we can do.

### 5.10 The Chinese restaurant process

This section follows Pitman's St Flour notes [3].

Suppose we have a consistent sequence of random permutations, i.e. $\sigma_n$ is a uniform permutation on $[n] = \{1, 2, \ldots, n\}$ for each $n$. Then we can write $\sigma_n$ as a product of its cycles.

For each $n$ we can get $\sigma_{n-1}$ by deleting the element $n$ from the cycle containing it in $\sigma_n$. For example, if $\sigma_5 = (123)(4)(5)$, then $\sigma_4 = (123)(4)$.

We now describe another process which gives rise to uniform random permutations. Suppose that there is an infinite number of tables in a restaurant and they are numbered $1, 2, \ldots$. Each table has infinite capacity. Customers numbered $1, 2, \ldots$ arrive into the restaurant. Customer 1 occupies table 1. Customer $n+1$ chooses with equal probability either to sit to the left of a customer already sitting in a table or to start a new table.

Hence the $n$-th person opens a new table with probability $1/n$ and sits to the left of $j \le n-1$ with probability $1/n$. So the probability he joins a group of size $n_i$ is $n_i/n$, which is the same as (5.8) and (5.9) for $\theta = 1$.

We define $\sigma_n : [n] \to [n]$ as follows. If customer $i$ is sitting to the left of customer $j$, then we set $\sigma_n(i) = j$, while if customer $i$ is sitting by himself, then we set $\sigma_n(i) = i$.

By induction, it is not hard to check that $\sigma_n$ has the properties listed above, i.e. for each fixed $n$ it is a random permutation and it also satisfies the consistency property.

Combining this with Ewens sampling formula from Theorem 5.19 we see that the total number of permutations of $\{1, \ldots, n\}$ with $a_1$ cycles of size 1, $a_2$ cycles of size 2 and so on, is equal to

$$\frac{n!}{\prod_{j=1}^{n} (j^{a_j} a_j!)}.$$

More properties of random permutations can be immediately read from this construction.

Let $K_n$ be the number of occupied tables when $n$ customers have arrived. Then this is equal to the number of cycles in the permutation $\sigma_n$. But

$$K_n = Z_1 + \ldots + Z_n,$$

where $Z_i$ is the indicator that the $i$-th customer occupies a new table. But this is a Bernoulli random variable of success probability $1/i$ and is independent of $Z_j$'s for $j \neq i$. Therefore,

$$\mathbb{E}[K_n] = \sum_{j=1}^{n} \frac{1}{j} \sim \log n \quad \text{as } n \to \infty.$$

Also we have almost surely

$$\frac{K_n}{\log n} \to 1 \quad \text{as } n \to \infty.$$

By the central limit theorem we obtain

$$\frac{K_n - \log n}{\sqrt{\log n}} \to \mathcal{N} \quad \text{as } n \to \infty \quad \text{in distribution,}$$

where $\mathcal{N}$ is a standard normal random variable.

Let $X_n$ be the indicator that the $(n+1)$-th customer sits at table 1. Then $X_n$ has the same dynamics as Pólya's urn. We now set $S_n = X_1 + \ldots + X_n$. In the proof of Theorem 5.16 we showed that $S_n$ is uniform on $\{1, \ldots, n\}$ and also that almost surely

$$\frac{S_n}{n} \to U \quad \text{as } n \to \infty,$$

where $U \sim U[0,1]$. In the language of permutations, $S_n + 1$ is the size of the cycle that contains the element 1. Therefore, we showed that the length of the cycle of the permutation containing the element 1 is uniform on $\{2, \ldots, n+1\}$. Since all elements of $\{1, \ldots, n\}$ are exchangeable for the permutation, the same is true for the cycle containing element $k \le n$.

Now, let's calculate the probability that two elements $i \neq j$ are in the same cycle of the permutation. By exchangeability again, this probability is the same as the probability that elements 1 and 2 are in the same cycle. Translating this question into the Chinese restaurant process, we see that this probability is equal to 1/2, since customer 2 must sit at table 1 which happens with probability 1/2.

Similarly the probability that $i, j, k$ are in the same cycle is equal to 1/3, since it is equal to

$$\mathbb{P}(\text{customers } 1, 2, 3 \text{ sit at the same table}) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

and for any number $\ell \le n$ we get

$$\mathbb{P}(\text{customers } 1, \ldots, \ell \text{ sit at the same table}) = \frac{1}{\ell}.$$

For $\ell = n$ we recover $\mathbb{P}(\sigma_n \text{ has one cycle}) = 1/n$.

## References


[1] Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.

[2] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[3] J. Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.