## Contents

### 1 Queueing Theory

1.1 Introduction ........................................... 3  
1.2 Example: $M/M/1$ queue ................................. 3  
1.3 Example: $M/M/\infty$ queue .......................... 4  
1.4 Burke’s theorem .......................................... 5  
1.5 Queues in tandem ....................................... 5  
1.6 Jackson Networks ....................................... 6  
1.7 Non-Markov queues: the $M/G/1$ queue. ............... 10

### 2 Renewal Theory

2.1 Introduction ........................................... 13  
2.2 Elementary renewal theorem ............................ 13  
2.3 Size biased picking ...................................... 14  
2.4 Equilibrium theory of renewal processes ............... 14  
2.5 Renewal-Reward processes .............................. 18  
2.6 Example: Alternating Renewal process ................. 18  
2.7 Example: busy periods of $M/G/1$ queue ............... 19  
2.8 $G/G/1$ queues and Little’s formula ................... 19

### 3 Population genetics

3.1 Introduction ........................................... 22  
3.2 Moran model ............................................. 22  
3.3 Fixation .................................................. 23  
3.4 The infinite sites model of mutations ................... 24  
3.5 Kingman’s $n$-coalescent ................................ 25  
3.6 Consistency and site frequency spectrum ............... 26  
3.7 Kingman’s infinite coalescent ......................... 26  
3.8 Intermezzo: Pólya’s urn and Hoppe’s urn* ............... 27  
3.9 Infinite Alleles Model .................................. 29  
3.10 Ewens sampling formula ............................... 29
1 Queueing Theory

1.1 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 notations 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 (typical $k = 1$ or $k = \infty$).

1.2 Example: $M/M/1$ queue

Customers arrive at rate $\lambda > 0$ and are served at rate $\mu > 0$ by a single server. 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

$$q_{i,i+1} = \lambda; q_{i,i-1} = \mu$$

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

**Theorem 1.1.** 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

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

If $W$ is the waiting time in the system of a customer before being served, then conditionally given $W > 0$,

$$W \sim \text{Exp}(\mu - \lambda).$$ (1)

**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. As concerns positive recurrence, observe that $\sup_i q_i < \infty$ so there is a.s. no explosion by a result from the lectures earlier. Hence we look for an invariant distribution. Furthermore, since $X$ is a birth and death chain, it suffices to solve the Detailed Balance Equations, which read:

$$\pi_n \lambda = \pi_{n+1} \mu$$

3
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 \( \pi_n \) is the distribution of a (shifted) geometric random variable. (Shifted because it can be equal to 0).

Suppose now that a customer arrives in the queue at some large time \( t \). Let \( N \) be the number of customers already in the queue at that time. We have \( W > 0 \) if and only if \( N \geq 1 \), and \( N \) has the distribution of a shifted geometric random variable. Conditionally on \( W > 0 \), \( N \) is thus an ordinary geometric random variable with parameter \( 1 - \rho \). Furthermore, if \( N = n \), then

\[
W = \sum_{i=1}^{n} T_i
\]

where \( T_i \) are iid exponential random variables with rate \( \mu \). We deduce from an exercise on Example Sheet 1 that (conditionally on \( W > 0 \)), \( W \) is thus an exponential random variable with parameter \( \mu(1 - \rho) = \mu - \lambda \).

**Example 1.2.** 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

\[
E(X) = \frac{1}{1 - \rho} - 1 = \frac{\rho}{1 - \rho} = \frac{\mu}{\mu - \lambda}.
\]

**1.3 Example: \( M/M/\infty \) 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 1.3.** \( X_t \) is a positive recurrent Markov chain for any \( \lambda, \mu > 0 \). Furthermore the Poisson \((\rho)\) distribution is invariant where \( \rho = \lambda/\mu \).

**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 total rate at which any one of them is being served is \( i\mu \), by superposition). Thus \( X \) is a birth and death chain; hence for an invariant distribution it suffices to solve the Detailed Balance Equations:

\[
\lambda \pi_{n-1} = n \mu \pi_n
\]

or

\[
\pi_n = \frac{1}{n!} \pi_0 \ldots = \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n \pi_0.
\]

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. We will show by hand that \( X \) is in fact recurrent. The idea is that for \( n \) sufficiently large, \( \mu n > \lambda \) so that \( X \) is 'dominated' by a random walk biased towards to 0, and is hence recurrent. More formally, let \( N \) be sufficiently large that \( \mu n \geq 2\lambda \) for all \( n \geq N \). If \( X \) is transient then we have

\[
P_{N+1}(X_t > N \text{ for all times } t) > 0
\]

But so long as \( X_t \geq N \), the jump chain \((Y_n)\) of \((X_t, t \geq 0)\) is such that \( P(Y_{n+1} = Y_n + 1) \leq 1/3 \) and \( P(Y_{n+1} = Y_n - 1) \geq 2/3 \). It follows that we can construct a \((2/3, 1/3)\) biased random walk \( \tilde{Y}_n \) such that \( Y_n \leq \tilde{Y}_n \) for all times \( n \). But \( \tilde{Y} \) is transient towards \(-\infty \) and hence is guaranteed
to return to $N$ eventually. This contradicts (2). Hence $X$ is recurrent and thus non-explosive. Since it has an invariant distribution we deduce that $X$ is also positive recurrent.

### 1.4 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 equilibrium. Let $D_t$ denote the number of customers who have departed the queue up to time $t$.

**Theorem 1.4.** (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 $(D_s, s \leq t)$.

**Remark 1.5.** 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 be equal to 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.

**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 $(\hat{X}_t, 0 \leq t \leq T)$ has the same distribution as $(X_t, 0 \leq t \leq T)$. 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 $T-t$. Since the time reversal of a Poisson process is a Poisson process, we deduce that $(D_t, t \leq T)$ is itself a Poisson process with rate $\lambda$.

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

**Remark 1.6.** The proof remains valid for any queue with birth and death queue length, at equilibrium, e.g. for a $M/M/\infty$ queue for arbitrary values of the parameters.

**Example 1.7.** 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 $60 \times 1 \times 10 = 600$.

### 1.5 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 1.8.** Let $X_t, Y_t$ denote the queue length in first (resp. second) queue. $(X,Y)$ 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 $1 - \rho_1, 1 - \rho_2$.

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

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

The rates are bounded so no explosion is possible. We can check by direct computation that $\pi Q = 0$ if and only $\pi$ has the desired form, hence the criterion for positive recurrence.

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 $\pi_1(m) = (1 - \rho_1)^m \rho_1^n$. Moreover we know by Burke’s theorem that (at equilibrium) the departure process is a Poisson process with rate $\lambda$. Hence 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 $\pi_2(n) = (1 - \rho_2)^n$. 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 distribution $\pi_1$ and $\pi_2$. Hence (since $(X, Y)$ is irreducible) it follows that this is the invariant distribution of $(X, Y)$. \qed

**Remark 1.9.** 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 1.10.** 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 $d \geq 3$.

### 1.6 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

$$
\begin{align*}
q(n, n + e_i) &= \lambda_i \\
q(n, n + e_j - e_i) &= \mu_i p_{ij} \text{ if } n_i \geq 1 \\
q(n, n - e_i) &= \mu_i p_{i0} \text{ if } n_i \geq 1.
\end{align*}
$$
What can be said about equilibrium, transience and recurrence? 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 1.11.** We say that a vector $(\bar{\lambda}_1, \ldots, \bar{\lambda}_N)$ satisfies the traffic equation if for all
$1 \leq i \leq N$:

$$\bar{\lambda}_i = \lambda_i + \sum_{j \neq i} \bar{\lambda}_j p_{ji} \quad (3)$$

The idea of (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 1.12.** There exists a unique solution, up to scaling, to the traffic equations (3).

*Proof.* Existence: Observe that the matrix $P = (p_{ij})_{0 \leq i,j \leq N}$ defines a stochastic matrix on
$\{0, \ldots, N\}$. The corresponding (discrete) Markov Chain is transient in the sense that it
is eventually absorbed at zero. Let $Z_n$ denotes a Markov chain in discrete time with this
transition matrix, started from the distribution $P(Z_0 = i) = \lambda_i / \lambda$, for $1 \leq i \leq N$, where
$\lambda = \sum_i \lambda_i$. Since $Z$ is transient, the number of visits $N_i$ to state $i$ by $Z$ satisfies $E(Z_i) < \infty$. But observe that

$$E(N_i) = P(Z_0 = i) + \sum_{n=0}^{\infty} P(Z_{n+1} = i)$$

$$= \frac{\lambda_i}{\lambda} + \sum_{n=0}^{\infty} \sum_{j=1}^{N} P(Z_n = j; Z_{n+1} = i)$$

$$= \frac{\lambda_i}{\lambda} + \sum_{n=0}^{\infty} \sum_{j=1}^{N} P(Z_n = j)p_{ji}$$

$$= \frac{\lambda_i}{\lambda} + \sum_{j=1}^{N} p_{ji} E(N_i).$$

Multiplying by $\lambda$, we see that if $\bar{\lambda}_i = \lambda E(N_i)$, then

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

which is the same thing as (3) as $p_{ii} = 0$.

Uniqueness: see example sheet.

We come to the main theorem of this section. This frequently appears in lists of the most
useful mathematical results for industry.
Theorem 1.13. (Jackson’s theorem, 1957). Assume that the traffic equations have a solution \( \bar{\lambda}_i \) such that \( \bar{\lambda}_i < \mu_i \) for every \( 1 \leq i \leq N \). Then the Jackson network is positive recurrent and

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

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 1.14. At equilibrium, the queue lengths \( X^i_t \) 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 nonreversible, 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 1.15. Suppose that \( X_t \) is a Markov chain on some state space \( S \), and that \( \pi(x) \geq 0 \) for \( x \in S \). Assume that for each \( x \in S \) we can find a partition of \( S \setminus \{x\} \), into say \( S^i_1, \ldots \) such that for all \( i \geq 1 \)

\[
\sum_{y \in S^i} \pi(x) q(x,y) = \sum_{y \in S^i} \pi(y) q(y,x).
\]

Then \( \pi \) is an invariant measure.

Definition 1.16. The equations (4) are called the Partial Balance Equations.

Proof. The assumptions say that for each \( x \) you can group the state 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,

\[
\pi(x) \sum_{y \neq x} q(x,y) = \sum_i \sum_{y \in S^i} \pi(x) q(x,y) = \sum_i \sum_{y \in S^i} \pi(y) q(y,x) = \sum_{y \neq x} \pi(y) q(y,x)
\]

so \( \pi Q(x) = 0 \) for all \( x \in S \).

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

\[
\tilde{q}(n,m) = \frac{\pi(m)}{\pi(n)} q(m,n).
\]

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
\[ A = \{e_i; 1 \leq i \leq N\}. \]

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

Let
\[ D_j = \{e_i - e_j; i \neq j\} \cup \{-e_j\}. \]

Thus if \( n \in S \) is a state and \( m \in D_j \) then \( n + m \) denotes any possible state after departure from a customer in queue \( j \).

We will show: for all \( n \in S \),
\[
\sum_{m \in D_j} q(n, m) = \sum_{m \in D_j} \tilde{q}(n, m) \tag{5}
\]
\[
\sum_{m \in A} q(n, m) = \sum_{m \in A} \tilde{q}(n, m) \tag{6}
\]

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

For the proof of (5), note that if \( m \in D_j \) 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 (5) is
\[
\sum_{m \in D_j} q(n, m) = \mu_j p_{j0} + \sum_{i \neq j} \mu_j p_{ji}
\]
\[ = \mu_j \]

which makes sense as services occur at rate \( \mu_j \).

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

Also,
\[
\tilde{q}(n, n - e_j) \frac{\pi(n - e_j)}{\pi(n)} q(n - e_j, n)
\]
\[ = \frac{\lambda_j}{\bar{\rho}_j}
\]

We deduce that the right hand side of (5) is given by
\[
\sum_{m \in D_j} \tilde{q}(n, m) = \frac{\lambda_j}{\bar{\rho}_j} + \sum_{i \neq j} \frac{\bar{\lambda}_i p_{ij}}{\bar{\rho}_j}
\]
\[ = \frac{\lambda_j}{\bar{\rho}_j} \quad \text{(by traffic equations)}
\]
\[ = \mu_j,
\]
as desired.

We now turn to (6). The left hand side is
\[ \sum_{m \in A} q(n, n + m) = \sum_i \lambda_i. \]

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 (6) is given by
\[ \sum_{m \in A} \tilde{q}(n, n + m) = \sum_i \bar{\lambda}_i p_{i0} \]
\[ = \sum_i \bar{\lambda}_i (1 - \sum_j p_{ij}) \]
\[ = \sum_i \bar{\lambda}_i - \sum_j \sum_i \bar{\lambda}_i p_{ij} \]
\[ = \sum_i \bar{\lambda}_i - \sum_j (\bar{\lambda}_j - \lambda_j) \quad \text{(by traffic equations)} \]
\[ = \sum_j \lambda_j, \]
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 of the queue, see Example Sheet 3.

\[ \square \]

1.7 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 1.17.** \((X(D_n), n \geq 1)\) forms a (discrete) Markov chain with transition probabilities given by
\[
\begin{pmatrix}
0 & p_0 & p_1 & p_2 & \ldots \\
p_0 & p_0 & p_1 & p_2 & \ldots \\
0 & 0 & p_0 & p_1 & p_2 & \ldots \\
\end{pmatrix}
\]
where for all \( k \geq 0 \), \( p_k = \mathbb{E}[\exp(-\lambda \xi)(\lambda \xi)^k/k!] \).
Remark 1.18. The form of the matrix is such that the first row is unusual. The other rows are given by the vector \((p_0, p_1, \ldots)\) which is pushed to the right at each row.

**Proof.** Assume \(X(D_n) > 0\). Then the \((n+1)\)th customer begins its 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

\[ P(A_n = k) = E(P(A_n = k | \xi_n)) = E[\exp(-\lambda \xi)(\lambda \xi)^k/k!] \]

in the statement. The result follows. \(\square\)

We write \(1/\mu = E(\xi)\), and call \(\rho = \lambda/\mu\) the traffic intensity. We deduce the following result:

**Theorem 1.19.** If \(\rho \leq 1\) 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.

**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 \(A_n - 1\). Hence it is transient if and only if \(E(A_n - 1) > 0\) or \(E(A_n) > 1\). But note that

\[ E(A_n) = E(E(A_n | \xi_n)) = \lambda E(\xi) = \rho \]

so the result follows.

The second proof consists in uncovering 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 \(E(A_n) \leq 1\) or \(\rho \leq 1\), since \(E(A_n) = \rho\). \(\square\)

As an application of the last argument we give the following example:

**Example 1.20.** The length of the busy period \(B\) of the \(M/G/1\) queue satisfies

\[ E(B) = \frac{1}{\mu - \lambda}. \]
To see this, adopt the branching process point of view. Let $A_1$ denote the number of offsprings 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

$$\mathbb{E}(B) = \mathbb{E}(\xi) + \mathbb{E}\left(\sum_{i=1}^{A_1} B_i | A_1, \xi_1\right)$$

$$= \mathbb{E}(\xi) + \mathbb{E}(A_1 \mathbb{E}(B))$$

$$= \mathbb{E}(\xi) + \rho \mathbb{E}(B)$$

hence

$$\mathbb{E}(B) = \frac{\mathbb{E}(\xi)}{1 - \rho}$$

and the result follows after some simplifications.

**Remark 1.21.** 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*. 
2 Renewal Theory

2.1 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, 10 minutes, and $\sqrt{\pi/10}$. This is illustrated by the following cases: if buses arrive exactly every 10 minutes, we probability 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 2.1. Let $(\xi_i, i \geq 1)$ be i.i.d. random variables with $\xi \geq 0$ and $\mathbb{P}(\xi > 0) > 0$. Let $T_n = \sum_{i=1}^{n} \xi_i$ and set

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

$(N_t, t \geq 0)$ 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 2.2. Since $\mathbb{P}(\xi > 0) > 0$ we have that $N_t < \infty$ a.s. Moreover one can see that $N_t \to \infty$ a.s.

2.2 Elementary renewal theorem

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

Theorem 2.3. If $1/\lambda = \mathbb{E}(\xi) < \infty$ then we have

$$\frac{N_t}{t} \to \lambda \text{ a.s.; and } \frac{\mathbb{E}(N_t)}{t} \to \lambda.$$

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

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

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

$$\frac{t}{N(t)} \to \frac{1}{\lambda}$$

a.s. and the result follows. \qed
2.3 Size biased picking

Suppose $X_1, \ldots, X_n$ are i.i.d. and positive. Let $S_i = X_1 + \ldots + X_i; 1 \leq i \leq n$. 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 $\hat{Y}$ denote the length of the interval containing $U$. What is the distribution of $\hat{Y}$? A first natural guess is that all the intervals are symmetric so we might guess $\hat{Y}$ has the same distribution as $Y = Y_1$, say. However this naïve 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, it can readily be checked that

$$P(\hat{Y} \in dy) = n y P(Y \in dy).$$

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 2.4.** Let $X$ be a nonnegative random variable with law $\mu$, and suppose $E(X) = m < \infty$. Then the size-biased distribution $\hat{\mu}$ is the probability distribution given by

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

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

**Remark 2.5.** Note that this definition makes sense because $\int_0^\infty \hat{\mu}(dy) = \int_0^\infty (y/m)\mu(dy) = m/m = 1$.

**Example 2.6.** If $X$ is uniform on $[0, 1]$ then $\hat{X}$ has the distribution $2xdx$ on $(0, 1)$. The factor $x$ biases towards larger values of $X$.

**Example 2.7.** If $X$ is an Exponential random variable with rate $\lambda$ then the size-biased distribution satisfies

$$P(\hat{X} \in dx) = \frac{x}{1/\lambda} \lambda e^{-\lambda x} dx = \lambda^2 x e^{-\lambda x} dx$$

so $\hat{X}$ is a Gamma $(2, \lambda)$ random variable. In particular $\hat{X}$ has the same distribution as the sum $X_1 + X_2$ of two independent Exponential random variables with rate $\lambda$.

2.4 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 2.8. 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 naïve 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.

Theorem 2.9. Let \((1/\lambda) = E(\xi)\). Then
\[ L(t) \to \hat{\xi} \] in distribution as \( t \to \infty \). Moreover, for all \( y \geq 0 \),
\[ P(E(t) \leq y) \to \lambda \int_{0}^{y} P(\xi > x)dx. \] as \( t \to \infty \) and the same result holds with \( A(t) \) in place of \( E(t) \). In fact,
\[ (L(t), E(t)) \to (\hat{\xi}, U\hat{\xi}) \] in distribution as \( t \to \infty \), where \( U \) is uniform on \((0, 1)\) and is independent from \( \hat{\xi} \). The same result holds with the pair \((L(t), A(t))\) instead of \((L(t), E(t))\).

Remark 2.10. One way to understand that theorem is that \( L(t) \) has the size-based distribution \( \xi \) and given \( L(t) \), the point \( t \) falls uniformly within the renewal interval of length \( L(t) \). That is the meaning of the uniform random variable in the limit (9). The restriction \( \xi \) continuous does not really need to be there.

Remark 2.11. Let us explain why (8) and (9) are consistent. Indeed, if \( U \) is uniform and \( \hat{\xi} \) has the size-biased distribution then
\[ P(U\hat{\xi} \leq y) = \int_{0}^{1} P(\hat{\xi} \leq y/u)du \]
\[ = \int_{0}^{1} (\int_{0}^{y/u} \lambda xP(\xi \in dx))du \]
\[ = \int_{0}^{\infty} \lambda xP(\xi \in dx) \int_{0}^{1} 1_{\{u \leq y/x\}}du \]
\[ = \int_{0}^{\infty} \lambda xP(\xi \in dx)(1 \wedge y/x) \]
\[ = \lambda \int_{0}^{\infty} (y \wedge x)P(\xi \in dx). \]
On the other hand,
\[
\lambda \int_0^y \mathbb{P}(\xi > z)\,dz = \lambda \int_0^y \int_z^\infty \mathbb{P}(\xi \in dx)\,dz = \lambda \int_0^\infty \mathbb{P}(\xi \in dx) \int_0^\infty 1_{\{z<y,z<x\}}\,dz = \lambda \int_0^\infty \mathbb{P}(\xi \in dx)(y \wedge x)
\]
so the random variable \( U_\xi \) indeed has the distribution function given by (8).

**Sketch of proof.** 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 (8) which is slightly easier. Consider the left-continuous excess \( E(t) = \lim_{s \to t^-} E(s) \) for \( t = 0, 1, \ldots \).

Then, as suggested by the picture in Figure 1, \( (E(t), t = 0, 1, \ldots) \) forms a discrete Markov chain with transitions
\[
p_{i,i-1} = 1
\]
for \( i \geq 1 \) and
\[
p_{0,n} = \mathbb{P}(\xi = n + 1)
\]
for all \( n \geq 1 \). 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
\[
\pi_n := \sum_{m \geq n+1} \mathbb{P}(\xi = m)
\]
is an invariant measure for this chain. This can be normalised to be a probability measure if \( E(\xi) < \infty \) in which case the invariant distribution is
\[
\pi_n = \lambda \mathbb{P}(\xi > n).
\]
We recognise the formula (8) in the discrete case where \( t \) and \( y \) are restricted to be integers.

We now consider the slightly more delicate result (9), still in the discrete case \( \xi \in \{1, 2, \ldots\} \). Of course, once this is proved, (7) 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

\[
S = \{(n, k) : 0 \leq k \leq n - 1\}.
\]

The transition probabilities are given by

\[
P(n, k) \rightarrow (n, k - 1) = 1
\]

if \( k \geq 1 \) and

\[
P(n, 0) \rightarrow (k, k - 1) = \mathbb{P}(\xi = k).
\]

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

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

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

**Example 2.12.** 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, \( \hat{\xi} = \text{Gamma}(2, \lambda) \) by Example 2.7. 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 2.13.** If \( \xi \) is uniform on \((0, 1)\) then for \( 0 \leq y \leq 1 \)

\[
\mathbb{P}(E_{\infty} \leq y) = \lambda \int_{0}^{y} \mathbb{P}(\xi > u)du = \lambda \int_{0}^{y} (1 - u)du = 2(y - y^2/2).
\]
2.5 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 2.14.** As \(t \to \infty\), if \(\mathbb{E}(|R|) < \infty\),

\[ \frac{R_t}{t} \to \lambda \mathbb{E}(R); \quad \text{and} \quad \frac{\mathbb{E}(R_t)}{t} \to \lambda \mathbb{E}(R). \]

Things are more interesting if we consider the current reward: i.e., \(r(t) = \mathbb{E}(R_{N(t)+1})\). 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 2.15.**

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

**Remark 2.16.** 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 \(1/\mathbb{E}(\xi)\).

2.6 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. 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 2.14 that if \(R_t\) is the total amount of time that the machine was on during \((0, t)\),

\[ \frac{R_t}{t} \to \frac{\mathbb{E}(X)}{\mathbb{E}(X) + \mathbb{E}(Y)}; \quad \text{and} \quad \frac{\mathbb{E}(R_t)}{t} \to \frac{\mathbb{E}(X)}{\mathbb{E}(X) + \mathbb{E}(Y)}. \] (10)

In reality there is a subtlety in deriving (10) from Proposition 2.14. 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 2.15.
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 (10) (since $\mathbb{E}(R_t) = \int_0^t p(s)ds$). Hence we deduce

$$p(t) \to \frac{\mathbb{E}(X)}{\mathbb{E}(X) + \mathbb{E}(Y)}$$

as $t \to \infty$.

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

Consider a $M/G/1$ 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 1.20)

$$\mathbb{E}(B) = \frac{1}{\mu - \lambda}.$$

We deduce that

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

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

### 2.8 $G/G/1$ queues and Little’s formula

Consider a $G/G/1$ queue. Let $A_n$ denote the intervals between arrival times of customers and $S_n$ their service times. It is not hard to prove the following result.

**Theorem 2.17.** 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, while if $\rho > 1$ then with positive probability it will never become empty.

**Sketch of 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 serviced up to time $t$). Then on $E$, we have

$$\frac{A_t}{t} \to \lambda; \frac{D_t}{t} \to \mu$$

by the elementary renewal theorem. Hence (on $E$), if $\rho > 1$, we have $D_t \gg A_t$ which is impossible. Hence $\mathbb{P}(E) = 0$ if $\rho < 1$. \hfill $\square$
We will now state and give the sketch of the proof of an interesting result concerning $G/G/1$ queues, known as Little’s formula. This relates the long-run queue length, waiting times of customers and arrival rates. To be precise, introduce the following quantities: let

$$L(t) = \lim_{t \to \infty} \frac{1}{t} \int_0^t X_s ds$$

where $X_s$ is the queue length at time $s$, and let

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

where $W_i$ is the waiting time including service of the $i$th customer. A priori it is not clear that these limits are well-defined.

**Theorem 2.18.** (Little’s formula). Assume $\rho < 1$. The limits defining $L$ and $W$ exist almost surely. Moreover,

$$L = \lambda W$$

where $1/\lambda = E(A_n)$.

**Sketch of proof.** Since $\rho < 1$, there are infinitely many times where the queue is idle and a new customer arrives. These instants, call them $T_n$, form a renewal process. Hence the limit for $L(t)$ comes from the renewal reward theory (more specifically Proposition 2.14), where the reward during $[T_n, T_{n+1}]$ corresponds to $R_n = \int_{T_n}^{T_{n+1}} X_s ds$.

Likewise the limit for existence of $W$ comes from the same result and a reward $R'_n$ given by the sum of waiting times of all customers arriving during $[T_n, T_{n+1}]$, and the elementary renewal theorem.

To prove the relation in the there, suppose that each customer pays £1 for each minute in the system (including when they are getting served). Hence at time $s$ the operator is collecting £$X_s$ per minute, and the total collected up to time $t$ is $\int_0^t X_s ds$. On the other hand, if all customers pay upfront when they enter the queue, then the amount collected is $\sum_{i=1}^{A_t} W_i$. We deduce:

$$\frac{1}{t} \int_0^t X_s ds = \frac{1}{t} \sum_{i=1}^{A_t} W_i + \text{error}$$

where error term comes from customers arriving before time $t$ and whose service ends after time $t$. But the right hand side can be rewritten as

$$\frac{1}{t} \sum_{i=1}^{A_t} W_i = \frac{A_t}{t} \frac{1}{A_t} \sum_{i=1}^{A_t} W_i \to \lambda W$$

as $t \to \infty$, while the left hand side converges to $L$ by definition. The result follows. \qed
Example 2.19. 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 (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 $\mu - \lambda$. 

3 Population genetics

3.1 Introduction

Sample the DNA of \( n \) individuals from a population. What patterns of diversity/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 3.1. The genome is the collection of all genetic information on 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... Different versions of the same gene are called alleles.

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.

3.2 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 3.2. Let \( N \geq 1 \). In the Moran model the population size is constant equal to \( N \). At rate 1, every individual dies. Simultaneously, a uniformly random chosen individual in the population gives birth.

Note: the population size stays constant through 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 let \( (N^i_j, t \geq 0) \) 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 3.3. The Moran model dynamics can be extended to \( t \in \mathbb{R} \), by using bi-infinite Poisson processes \( (N^i_j, t \in \mathbb{R}) \).
3.3 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 3.4.** 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,

$$\mathbb{E}(\tau | X_0 = i) = \frac{i}{q_i} \left( \sum_{j=1}^{N-i} \frac{N-j}{j} + \sum_{j=i}^{N} \frac{i}{j} \right).$$

**Remark 3.5.** 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

$$q_{i,i+1} = (N-i) \times i/N$$

(13)

(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,

$$q_{i,i-1} = i \times (N-i)/N.$$ (14)

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,

$$\mathbb{P}(T_N < T_0) = i/N.$$ (15)

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 decomposing on the first step, the probability to not return to $j$ is given by

$$\mathbb{P}_j(\text{ no return to } j) = \frac{1}{2j} + \frac{1}{2} \frac{1}{N-j},$$

by using (15) on the interval $[0,j]$ and $[j,N]$ respectively. Hence

$$\mathbb{E}_j(\tau_j) = \frac{1}{q_j} \frac{2j(N-j)}{N} = 1.$$

Consequently,

$$\mathbb{E}_i(\tau) = \sum_j \mathbb{E}_j(\tau_j) = \sum_i \mathbb{P}_i(X_t = j \text{ for some } t \geq 0) \times \mathbb{E}_j(\tau_j) = \sum_{j \geq i} \frac{i}{j} + \sum_{j < i} \frac{N-i}{N-j}.$$

The result follows. □
3.4 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). Assuming that no two mutations affect the same site, it 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? We call \( M_j(n) = \# \) alleles carried by exactly \( j \) individuals in the sample.

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). Thus, if we are given a table with the DNA sequences of all \( n \) individuals in the sample, \( M_j(n) \) will be the number of sites (bases) where exactly \( j \) individuals carry a base which differs from everyone else in the sample.

**Example 3.6.** Suppose the DNA sequences in a sample are as follows

1: \( \ldots A T T T C G G G T C \ldots \)
2: \( \ldots - A - G - - - - - - - \ldots \)
3: \( \ldots - - - - - - - C - \ldots \)
4: \( \ldots - - - G - - - - C - \ldots \)
5: \( \ldots - A - - - - - - - \ldots \)
6: \( \ldots - - - - - - - C - \ldots \)
7: \( \ldots - - - - - - - C - \ldots \)

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 \( M_3(n) = 1 \) (last but one).

Our first result tells us what happens in the (unrealistic) case where \( n = N \).

**Theorem 3.7.** Let \( \theta = uN \). Then

\[ \mathbb{E}(M_j(N)) = \frac{\theta}{j}. \]

**Proof.** Mutations occur at a total rate of \( u \times N = \theta \) over 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, \( Y \) evolves like the Markov chain in the previous theorem, i.e., has the \( Q \)-matrix given by (13) and (14). 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

\[
\mathbb{E}(M_j(N)) = \int_0^\infty uN dt p_t(1,j) = \theta \mathbb{E}_1(\tau_j) = \theta \mathbb{P}_1(X_t = j \text{ for some } t \geq 0) \mathbb{E}_j(\tau_j) = \theta \times (1/j) \times 1,
\]

as desired.

\[ \square \]

### 3.5 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 0. As \( -t \) goes further and further back in time, it may happen that the ancestor for two individuals in the population become 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 \( \{1, \ldots, n\} \)) 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 3.8.** \((\Pi_{Nt/2}, t \geq 0)\) is a Markov chain in \( \mathcal{P}_n \) with

\[
q_{\pi,\pi'} = \begin{cases} 
1 & \text{if } \pi' \text{ can be obtained from } \pi \text{ by coagulating two of its blocks} \\
-(\binom{k}{2}) & \text{if } \pi' = \pi \\
0 & \text{else} .
\end{cases}
\]

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 rate 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 number of Kingman’s \( n \)-coalescent 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 \( K_t \).
3.6 Consistency and site frequency spectrum

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 3.9.** Let \( \Pi^n \) be Kingman's \( n \)-coalescent. Then \( \Pi^n|_{[n-1]} \) 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.

The sampling consistency has some fundamental consequences. For instance, recall the infinite sites model, where each individual is subject to mutation at rate \( u > 0 \), independently, and all mutations are distinct. Recall the quantity \( M_j(n) \), the number of distinct alleles carried by exactly \( j \) individuals. 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. We have shown in Theorem 3.7 that for such a tree \( \mathbb{E}(M_j(N)) = \theta/j \) where \( \theta = uN \).

By scaling, \( M_j(N) \) would be unchanged if we consider a Kingman \( N \)-coalescent tree on which mutations fall at rate \( \theta/2 \) per branch and per unit length. But this is a question for which the only parameter is the number \( N \) of initial branches of the tree. We deduce the following result:

**Theorem 3.10.** 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 3.11.** 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. E.g in the example a few pages before, \( S_n = 3 \). Then we deduce from the above theorem that

\[
\mathbb{E}(S_n) = \theta(1 + \ldots + \frac{1}{n}) \sim \theta \log n.
\]

as \( n \to \infty \).

3.7 Kingman’s infinite coalescent

The sampling consistency can be used to 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 \( n \geq 1 \), \( \Pi|_{[n]} \) has the law of Kingman’s \( n \)-coalescent.

**Definition 3.12.** \((\Pi_t, t \geq 0)\) is called Kingman’s infinite coalescent.

Initially we have \( \Pi_0 \) consisting of infinitely many singletons. How does it look like for positive times? E.g will it ever completely coalesce? One remarkable phenomenon with Kingman’s coalescent is the following fact.
Theorem 3.13. 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. By results in measure theory,
\[
P(|\Pi_t| \geq M) = \lim_{n \to \infty} P(|\Pi^n_t| \geq M)
\]
\[
= P\left( \sum_{j=M+1}^{\infty} \tau_j \geq t \right)
\]
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 $\left( \frac{j}{2} \right)$. By Markov’s inequality
\[
P(|\Pi_t| \geq M) \leq \frac{E(\sum_{j=M+1}^{\infty} \tau_j)}{t} \leq \frac{1}{t} \sum_{j=M+1}^{\infty} \frac{1}{j} = \frac{2}{t} \sum_{j=M+1}^{\infty} \frac{1}{j(j-1)}
\]
This tends to 0 as $M \to \infty$, so the result follows. \qed

3.8 Intermezzo: Pólya’s urn and Hoppe’s urn

The proofs in this section is 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 3.14. 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 making a few simple computation. What is the probability to get first $m$ black balls and then $n - m$ balls (in that order)? We see that it is
\[
\frac{1}{2} \frac{2}{3} \cdots \frac{m-1}{m} \times \frac{1}{m+1} \cdots \frac{n-m}{n+1} = \frac{(m-1)!(n-m)!}{(n+1)!}.
\]
The first factor accounts 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 to 1 \(m-1\) and from 1 to \(n\) respectively, albeit at different times than in the above order. Still the product is unchanged. Hence

\[
P(X_{n+2} = m) = \binom{n}{m} \frac{(m-1)!(n-m)!}{(n+1)!} \frac{n!}{m!(n-m)!} = \frac{1}{n+1},
\]

so \(X_{n+2}\) is uniformly distributed over \(\{1, \ldots, n+1\}\). 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 \((m-1)!(n-m)!/(n+1)!\). Clearly the probability is also invariant under permutation of the sequence, so these two processes must be identical! It now follows that by the law of large numbers, \(X_n/n \rightarrow U\), almost surely.

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 the 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. It can be deduced from Pólya’s theorem above that there will be ultimately infinitely many colours in the urn. The proportion of balls of a given colour \(i \geq 1\) converges a.s. to a random variable \(p_i\), and \(\sum_i p_i = 1\) a.s.
3.9 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 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 \( \{1, \ldots, n\} \)) 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 3.15.** \( \Pi_n \) 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.

**Example 3.16.** Suppose \( n = 8 \) and the eye colour of the sample is

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\text{blue} & \text{red} & \text{brown} & \text{green} & \text{brown} & \text{yellow} & \text{red} & \text{brown}
\end{array}
\]

Then \( \Pi_n \) has 4 blocks (corresponding to the five colours in the sample): \( \{1\}, \{2, 7\}, \{3, 5, 8\}, \{4\}, \{6\} \).

Hence \( a_1 = 3, a_2 = 1, a_3 = 1 \).

3.10 Ewens sampling formula

It turns out that we can describe the distribution of \( \Pi_n \) or, equivalently, of \( (A_1, \ldots, A_n) \), 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 3.17.** *(Ewens sampling formula)* Let \( a_j \) be such that \( \sum_{j=1}^n ja_j = n \). Then

\[
P(A_1 = a_1, \ldots, A_n = a_n) = \frac{n!}{\theta(\theta+1) \ldots (\theta+n-1)} \prod_{j=1}^n \frac{(\theta/j)^{a_j}}{a_j!}.
\] (17)

**Remark 3.18.** It is far from obvious that the right hand side adds up to one!

**Remark 3.19.** As we will see, an equivalent way of stating the formula is that

\[
P(\Pi_n = \pi) = \frac{\theta^k}{\theta(\theta+1) \ldots (\theta+n-1)} \prod_{i=1}^k (n_i - 1)!
\] (18)
Figure 2: 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 $T_1, \ldots, T_{n-1}$ in the proof.

where $k$ is the number of $\pi$, and $\pi \in \mathcal{P}_n$ is arbitrary. This is the version which we will prove below.

Remark 3.20. We see from (17) that the distribution of $(A_1, \ldots, A_n)$ is that of independent Poisson random variables with mean $\theta/j$ conditioned so that $\sum_j j A_j = n$. 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 $\theta/j$. E.g., we have (for large $n$) $\Pr(A_j = 0) \approx e^{-\theta/j}$.

Proof. 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 $T_{n-1}, \ldots, T_1$ 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 $T_{n-1}$. Between times $T_{m-1}$ and $T_m$ there are $m$ branches. These branches are already grouped into blocks sharing the same allelic type. At time $T_m$ 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, of 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”: during $T_m$ and $T_{m+1}$ there were $m + 1$ branches. So there are $m + 1$ exponential clocks with rate $\theta/2$ corresponding to mutation, and $m(m + 1)/2$ clocks with rate 1, corresponding with coalescence.

Hence the probability to form a new allelic group for the new branch at time $T_m$ is

$$\mathbb{P}(\text{new block}) = \frac{(m + 1)\theta/2}{(m + 1)\theta/2 + m(m + 1)/2} = \frac{\theta}{\theta + m}.$$  

The probability to join an existing group of size $n_i$ is

$$\mathbb{P}(\text{join a group of size } n_i) = \frac{m(m + 1)/2}{m(m + 1)/2} \times \frac{n_i}{m} = \frac{n_i}{m + \theta}.$$  

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 now can prove

$$\mathbb{P}(\Pi_n = \pi) = \frac{\theta^k}{\theta(\theta + 1)\ldots(\theta + n - 1)} \prod_{i=1}^{k} (n_i - 1)!$$

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

$$\mathbb{P}(\Pi_n = \pi) = \mathbb{P}(\Pi_{n-1} = \pi') \times \frac{\theta}{\theta + n - 1}$$

$$= \frac{\theta^{k-1}}{\theta(\theta + 1)\ldots(\theta + n - 2)} \prod_{i=1}^{k-1} (n_i - 1)! \times \frac{\theta}{\theta + n - 1}$$

$$= \frac{\theta^k}{\theta(\theta + 1)\ldots(\theta + n - 1)} \prod_{i=1}^{k} (n_i - 1)!$$

as desired.

In case (b),

$$\mathbb{P}(\Pi_n = \pi) = \mathbb{P}(\Pi_{n-1} = \pi') \times \frac{n_j - 1}{\theta + n - 1}$$

$$= \frac{\theta^k}{\theta(\theta + 1)\ldots(\theta + n - 2)} \prod_{i=1; i \neq j}^{k} (n_i - 1)! \times (n_j - 2)! \times \frac{n_j - 1}{\theta + n - 1}$$

$$= \frac{\theta^k}{\theta(\theta + 1)\ldots(\theta + n - 1)} \prod_{i=1}^{k} (n_i - 1)!$$

as desired. Either way, the formula is proved.
Step 3. Combinatorics: we show that the two formulas (18) and (17) 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 $\sum_j j a_j = n$ and $\pi$ is a given partition having $(a_j)$ as its allele count, we have:

$$P(A_1 = a_1, \ldots, A_n = a_n) = P(\Pi_n = \pi) \times \# \{\text{partitions with this allele count}\}$$

$$= \frac{\theta^k}{\theta(\theta + 1) \ldots (\theta + n - 1)} \prod_{j=1}^{k} (n_j - 1)! \times \frac{1}{n!} \prod_{j=1}^{k} \frac{1}{n_j!} \prod_{i=1}^{n} a_i!$$

$= \frac{\theta^k n!}{\theta(\theta + 1) \ldots (\theta + n - 1)} \times \prod_{j=1}^{k} \frac{1}{n_j!} \prod_{i=1}^{n} a_i!$

$= \frac{n!}{\theta(\theta + 1) \ldots (\theta + n - 1)} \prod_{i=1}^{n} \frac{(\theta/j)^{a_j}}{a_j!}$,

as desired. \hfill \Box

**Corollary 3.21.** Let $K_n = \# \text{ distinct alleles in a sample of size } n$. Then

$$\mathbb{E}(K_n) = \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} \sim \theta \log n;$$

$$\text{var}(K_n) \sim \theta \log n$$

and

$$\frac{K_n - \mathbb{E}(K_n)}{\sqrt{\text{var}(K_n)}} \rightarrow \mathcal{N}(0, 1)$$

a standard normal random variable.

**Proof.** This follows from the Hoppe urn representation in the proof. At each step, 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 $\mathbb{E}(K_n)$, $\text{var}(K_n)$ follow, and the central limit theorem comes from computing the characteristic function. \hfill \Box

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.