Lecture 5

Continuous-time Markov chains

Reading: Norris 2.1, 2.6
Further reading: Grimmett-Stirzaker 6.9; Ross 6.1-6.3; Norris 2.9

In this lecture, we generalise the notion of a birth process to allow deaths and other transitions, actually transitions between any two states in a state space $\mathbb{S}$, just as for discrete-time Markov chains and using the notion of a discrete-time Markov chain.

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

5.1 Definition and terminology

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

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

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

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

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

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

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

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

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

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

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

Example 25 (continued) For birth processes, we obtain

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

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

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

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

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

5.2 Construction

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

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

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

and

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

Then

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

is a (ν, Q)*-Markov chain.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 30 (M/M/s queue) If there are $s \geq 1$ servers in the system, the rate at which customers leave is s-fold, provided there are at least s customers in the system.

We obtain the Q-matrix

$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \mu & -\mu-\lambda & \lambda & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 2\mu & -2\mu-\lambda & \lambda & \ddots & 0 & 0 & 0 & \ddots \\ 0 & 0 & 3\mu & -3\mu-\lambda & \ddots & 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \ddots & -s\mu-\lambda & \lambda & 0 & \ddots \\ 0 & 0 & 0 & 0 & \ddots & s\mu & -s\mu-\lambda & \lambda & \ddots \\ 0 & 0 & 0 & 0 & \ddots & 0 & s\mu & -s\mu-\lambda & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

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

Lecture 6

Transition semigroups

Reading: Norris 2.8, 3.1
Further reading: Grimmett-Stirzaker 6.8 (12)-(17), 6.9; Ross 6.4; Norris 2.7, 2.10

In this lecture we establish transition matrices $P(t)$, $t \geq 0$, for continuous-time Markov chains. This family of matrices are the analogues of n-step transition matrices $\Pi^n = (\pi_{ij}^{(n)})_{i,j \in \mathbb{S}}$, $n \geq 0$, for discrete-time Markov chains. While we will continue to use the Q-matrix to specify the distribution of a continuous-time Markov chain, transition matrices $P(t)$ give some of the most important probabilities related to a continuous-time Markov chain, but they are available explicitly only in a limited range of examples.

6.1 The semigroup property of transition matrices

As a consequence of the Markov property of continuous-time Markov chains, the probabilities $\mathbb{P}(X_{t+s} = j | X_t = i)$ do not depend on t. We denote by

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

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

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

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

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

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

Proof: Just note that for all $i, k \in \mathbb{S}$

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

where we applied the Markov property. □

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

6.2 Backward equations

The following result is useful to calculate transition probabilities.

Proposition 33 *The transition matrices $(P(t))_{t\geq 0}$ of a minimal (ν, Q)-Markov chain satisfy the backward equation*

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

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

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

Proof: We first show that $(P(t))_{t\geq 0}$ solves $P'(t) = QP(t)$, i.e. for all $i, k \in \mathbb{S}$, $t \geq 0$

$$p'_{ik}(t) = \sum_{j\in\mathbb{S}} q_{ij} p_{jk}(t).$$

We start by a one-step analysis (using the strong Markov property at the first jump time T_1, or directly identifying the structure of the post-T_1 process) to get

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

i.e.

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

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

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

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

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

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

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

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

then as above

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

and therefore

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

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

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

6.3 Forward equations

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

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

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

Proof: See Assignment question A.3.5. □

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

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

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

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

Remark 36 Transition semigroups and the Markov property can form the basis for a definition of continuous-time Markov chains. In order to match our definition, we could say that a (ν, Q)-Markov chain is a process with *right-continuous sample paths* in $\mathbb{S}$ such that

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = p_{i_n, i_{n+1}}(t_{n+1} - t_n)$$

for all $0 \le t_0 \le t_1 \le \ldots \le t_{n+1}$ and $i_0, \ldots, i_{n+1} \in \mathbb{S}$, where $P(t)$ satisfies the forward equations. See Norris 2.8.

6.4 Example

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

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

and it is easily seen inductively (fix i and proceed $n = 0, 1, 2, \ldots$) that Poisson probabilities

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

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

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

and solve inductively (fix j and proceed $i = j, j-1, \ldots, 0$). We have seen an easier way to derive the Poisson transition probabilities in Remark 2. The link between the two ways is revealed by the passage to probability generating functions

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

which then have to satisfy differential equations

$$\frac{\partial}{\partial t} G_i(z, t) = \sum_{n=0}^{\infty} z^{i+n} p'_{i,i+n}(t) = \lambda(z - 1) G_i(z, t), \qquad G_i(z, 0) = \mathbb{E}_i(z^{X_0}) = z^i.$$

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

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

that yield here

$$m_i'(t) = \frac{\partial}{\partial z}\frac{\partial}{\partial t}G_i(z,t)\bigg|_{z=1-} = \lambda\, G_i(z,t)\big|_{z=1-} = \lambda, \qquad m_i(0) = \mathbb{E}_i(X_0) = i.$$

Often (even in this case), this can be solved more easily than the differential equation for probability generating functions. Together with a similar equation for the variance, we can capture the two most important distributional features of a model.

6.5 Matrix exponentials

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

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

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

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

Lecture 7

The class structure of continuous-time Markov chains

Reading: Norris 3.2-3.5
Further reading: Grimmett-Stirzaker 6.9

In this lecture, we introduce for continuous-time chains the notions of irreducibility and positive recurrence that will be needed for the convergence theorems in Lecture 8.

7.1 Communicating classes and irreducibility

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

Definition 38 Let X be a continuous-time Markov chain.

(a) We say that $i \in \mathbb{S}$ *leads to* $j \in \mathbb{S}$ and write $i \rightarrow j$ if

$$\mathbb{P}_i(X_t = j \text{ for some } t \geq 0) = \mathbb{P}_i(T_{\{j\}} < \infty) > 0, \quad \text{where } T_{\{j\}} = \inf\{t \geq 0 : X_t = j\}.$$

(b) We say i *communicates with* j and write $i \leftrightarrow j$ if both $i \rightarrow j$ and $j \rightarrow i$.

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

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

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

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

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

Proposition 39 *Let X be a minimal continuous-time Markov chain. For $i, j \in \mathbb{S}$, $i \neq j$, the following are equivalent*

(i) $i \rightarrow j$ *for* X.

(ii) $i \rightarrow j$ *for* M.

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

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

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

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

(ii)⇒(iii): From the discrete-time theory, we know that $i \rightarrow j$ for M implies that there is a path $(i_0, \ldots, i_n)$ from i to j with

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

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

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

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

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

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

for all $t > 0$. For the last equality, we used the Markov property which implies that for all $m = 1, \ldots, n$

$$\begin{aligned} \mathbb{P}(X_{mt/n} = i_m | X_{kt/n} = i_k \text{ for all } k = 0, \ldots, m-1) &= \mathbb{P}(X_{mt/n} = i_m | X_{(m-1)t/n} = i_{m-1}) \\ &= p_{i_{m-1}, i_m}(t/n). \end{aligned}$$

□

Condition (iv) shows that the situation is simpler than in discrete-time where it may be possible to reach a state, but only after a certain length of time, and then only periodically.

7.2 Recurrence and transience

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

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

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

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

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

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

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

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

the first passage time of X to state i. Note that we require the chain to do at least one jump. This is to force X to leave i first if $X_0 = i$. We also define the successive passage times by $N_i^{(1)} = N_i$ and $N_i^{(m+1)} = \inf\{n > N_i^{(m)} : M_n = i\}$, $m \geq 1$, for M, and

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

$m \geq 1$, for X.

Proposition 41 *$i \in \mathbb{S}$ is recurrent (transient) for a minimal continuous-time Markov chain X if and only if i is recurrent (transient) for the jump chain M.*

Proof: Suppose, i is recurrent for the jump chain M, i.e. M visits i infinitely often, at steps $(N_i^{(m)})_{m\geq 1}$. If we denote by $1_{\{X_0=i\}}$ the random variable that is 1 if $X_0 = i$ and 0 otherwise, the total amount of time that X spends at i is

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

with probability 1 by the argument for Proposition 16 (convergent and divergent sums of independent exponential variables) since $Z_{N_m} \sim Exp(\lambda_i)$ and the sum of their (identical!) inverse parameters is infinite. In particular $\{t \geq 0 : X_t = i\}$ must be unbounded with probability 1.

Suppose, i is transient for the jump chain M, then there is a last step $L < \infty$ away from i and

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

is bounded with probability 1.

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

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

Corollary 42 *Every state $i \in \mathbb{S}$ is either recurrent or transient for X.*

Recall that a class property is a property of states that either all states in a (communicating) class have or all states in a (communicating) class don't have.

Corollary 43 *Recurrence and transience are class properties.*

Proof: If i is recurrent and $i \leftrightarrow j$, for X, then i is recurrent and $i \leftrightarrow j$ for M. From discrete-time Markov chain theory, we know that j is recurrent for M. Therefore j is recurrent for X.

The proof for transience is similar. □

Proposition 44 *For any $i \in \mathbb{S}$ the following are equivalent:*

(i) i *is recurrent for* X.

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

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

Proof: (iii)⇒(ii): One can deduce this from the corresponding discrete-time result, but we give a direct argument here. Assume $\lambda_i > 0$ and $h_i = \mathbb{P}_i(H_i = \infty) > 0$. Then, the strong Markov property at $H_i^{(m)}$ states that, given $H_i^{(m)} < \infty$, the post-$H_i^{(m)}$ process $X^{(m+1)} = (X_{H_i^{(m)}+t})_{t\ge 0}$ is distributed as X and independent of the pre-$H_i^{(m)}$ process. Now the total number G of visits of X to i must have a geometric distribution with parameter h_i since $\mathbb{P}_i(G = 1) = h_i$ and $\mathbb{P}_i(G = m|G \ge m) = h_i$, $m \ge 2$. Therefore, the total time spent in i is

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

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

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

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

7.3 Positive and null recurrence

As in the discrete-time case, there is a link between recurrence and the existence of invariant distributions. More precisely, recurrence is strictly weaker. The stronger notion required is positive recurrence:

Definition 45 A state $i \in \mathbb{S}$ is called *positive recurrent* if either $\lambda_i = 0$ or $m_i = \mathbb{E}_i(H_i) < \infty$. Otherwise, we call i null recurrent.

Fact 46 *Positive recurrence is a class property.*

7.4 Examples

Example 47 The M/M/1 queue with $\lambda > 0$ and $\mu > 0$ is irreducible since for all $m > n \geq 0$, we have $q_{m,m-1} \dots q_{n+1,n} = \mu^{m-n} > 0$ and $q_{n,n+1} \dots q_{m-1,m} = \lambda^{m-n} > 0$ and Proposition 39 yields $m \leftrightarrow n$.

$\lambda > \mu$ means that customers arrive at a higher rate than they leave. Intuitively, this means that $X_t \to \infty$ (this can be shown by comparison of the jump chain with a simple random walk with up probability $\lambda/(\lambda + \mu) > 1/2$). As a consequence, $L_i = \sup\{t \geq 0 : X_t = i\} < \infty$ for all $i \in \mathbb{N}$, and since $\{t \geq 0 : X_t = i\} \subset [0, L_i]$, we deduce that i is transient.

$\lambda < \mu$ means that customers arrive at a slower rate than they can leave. Intuitively, this means that X_t will return to zero infinitely often. The mean of the return time can be estimated by comparison of the jump chain with a simple random walk with up probability $\lambda/(\lambda + \mu) < 1/2$:

$$\begin{aligned}
\mathbb{E}_0(H_0) &= \mathbb{E}\left(\sum_{k=0}^{N_0} Z_k\right) = \sum_{n=2}^{\infty} \mathbb{P}(N_0 = n)\left(\sum_{k=0}^{n} Z_k \middle| N_0 = n\right) \\
&= \frac{1}{\lambda} + \sum_{n=2}^{\infty} \mathbb{P}(N_0 = n)\left(\sum_{k=1}^{n} Y_k\right) = \frac{1}{\lambda} + \sum_{n=2}^{\infty} \mathbb{P}(N_0 = n)\frac{n}{\lambda+\mu} = \frac{1}{\lambda} + \frac{\mathbb{E}(N_0)}{\lambda+\mu} < \infty.
\end{aligned}$$

Therefore, 0 is positive recurrent. Since positive recurrence is a class property, all states are positive recurrent.

For $\lambda = \mu$, the same argument shows that 0 is null-recurrent, by comparison with simple symmetric random walk.

Note in each case, that the jump chain is not a simple random walk, but coincides with a simple random walk until it hits zero. This is enough to calculate $\mathbb{E}_0(N_0)$.

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

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

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

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

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

Lecture 8

Convergence to equilibrium

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

In Lecture 7 we studied the class structure of continuous-time Markov chains. We can summarize the findings by saying that the state space can be decomposed into (disjoint) communicating classes

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

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

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

8.1 Invariant distributions

Note that for an initial distribution ν on $\mathbb{S}$, $X_0 \sim \nu$ we have

$$\mathbb{P}(X_t = j) = \sum_{i\in\mathbb{S}} \mathbb{P}(X_0 = i)\mathbb{P}(X_t = j | X_0 = i) = (\nu P(t))_j$$

where $\nu P(t)$ is the product of a row vector ν with the matrix $P(t)$, and we extract the jth component of the resulting row vector.

Definition 49 A distribution ξ on $\mathbb{S}$ is called invariant for a continuous-time Markov chain if $\xi P(t) = \xi$ for all $t \geq 0$.

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

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

Proof: If $\xi P(t) = \xi$ for all $t \geq 0$, then by the forward equation

$$\xi Q = \xi P(t) Q = \xi P'(t) = \xi \lim_{h \to 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \to 0} \frac{\xi P(t+h) - \xi P(t)}{h} = 0.$$

If $\xi Q = 0$, we have

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

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

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

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

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

Fact 52 *An irreducible (minimal) continuous-time Markov chain is positive recurrent if and only if it has an invariant distribution. An invariant distribution ξ can then be given by*

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

where $m_i = \mathbb{E}_i(H_i)$ is the mean return time to i and $\lambda_i = -q_{ii}$ the holding rate in i.

The proof is quite technical and does not give further intuition. The analogous result for discrete chains holds and gives $\eta_i = 1/\mathbb{E}_i(N_i)$ as invariant distribution. The further factor λ_i occurs because a chain in stationarity is likely to be found in i if the return time is short and the holding time is long; both observations are reflected through the inverse proportionality to $m_i$and λ_i, respectively. Since this is a key result for both Convergence Theorem and Ergodic Theorem, the diligent reader may want to refer to Norris Theorem 3.5.3.

Example 53 Consider the M/M/1 queue of Example 29. The equations $\xi Q = 0$ are given by

$$-\lambda\xi_0 + \mu\xi_1 = 0, \quad \lambda\xi_{i-1} - (\lambda+\mu)\xi_i + \mu\xi_{i+1} = 0, \quad i \geq 1.$$

This system of linear equations (for the unknowns ξ_i, $i \in \mathbb{N}$) has a probability mass function as its solution if and only if $\lambda < \mu$. It is given by the geometric probabilities

$$\xi_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right), \quad i \in \mathbb{N}.$$

By Fact 52, we can calculate $\mathbb{E}_i(H_i) = m_i = 1/(\lambda_i\xi_i)$. In particular, for $i = 0$, we have the length of a full cycle beginning and ending with an empty queue. Since the initial empty period has average length $1/\lambda$, the busy period has length

$$\mathbb{E}_0(H_0) - 1/\lambda = \frac{1}{\lambda\,(1-\lambda/\mu)} = \frac{1}{\mu-\lambda}.$$

Note that this tends to infinity as $\lambda \uparrow \mu$.

8.2 Convergence to equilibrium

The convergence theorem is of central importance in applications since is it often assumed that a system is in equilibrium. The convergence theorem is a justification for this assumption, since it means that a system must only be running long enough to be (approximately) in equilibrium.

Proposition 54 *Let $X = (X_t)_{t\geq 0}$ be a (minimal) irreducible positive-recurrent continuous-time Markov chain, $X_0 \sim \nu$, and ξ an invariant distribution, then*

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

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

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

Sketch of proof: Let X be the continuous-time Markov chain starting according to ν, Y an independent continuous-time Markov chain with the same Q-matrix, but starting from the invariant distribution ξ. Choose $i \in \mathbb{S}$ and define $T = \inf\{t \geq 0 : (X_t, Y_t) = (i,i)\}$ the time they first meet (in i, to simplify the argument). A third process is constructed $\widetilde{X}_t = X_t$, $t < T$ (following the ν-chain before T), $\widetilde{X}_t = Y_t$, $t \geq T$ (following the ξ-chain after T). The following three steps complete the proof:

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

2. the third chain $\widetilde{X}$ has the same distribution as the ν-chain X;
3. the third chain (which eventually coincides with the ξ-chain Y) is asymptotically in equilibrium in the sense of the convergence statement in Fact 54.

□

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

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

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

Proof: A proof using renewal theory is in assignment question A.5.4. □

We interpret this as follows. For any initial distribution, the long-term proportions of time spent in any state i approaches the invariant probability for this state. This result establishes a time-average analogue for the spatial average of Fact 54. This is of great practical importance, since it allows us to *observe* the invariant distribution by looking at time proportions over a long period of time. If we tried to observe the stationary distribution using Fact 54, we would need many independent observations of the same system at a large time t to estimate ξ.

8.3 Detailed balance equations and time reversal

Proposition 56 *Consider a Q-matrix Q. If the* detailed balance equations

$$\xi_i q_{ij} = \xi_j q_{ji}, \qquad i, j \in \mathbb{S},$$

have a solution $\xi = (\xi_i)_{i\in\mathbb{S}}$, then ξ is a stationary distribution.

Proof: Let ξ be such that all detailed balance equations hold. Then fix $j \in \mathbb{S}$ and sum the equations over $i \in \mathbb{S}$ to get

$$(\xi Q)_j = \sum_{i\in\mathbb{S}} \xi_i q_{ij} = \sum_{i\in\mathbb{S}} \xi_j q_{ji} = \xi_j \sum_{i\in\mathbb{S}} q_{ji} = 0$$

since the row sums of any Q-matrix vanish (see Remark 26). Therefore $\xi Q = 0$, as required. □

Note that (in the case of finite $\#\mathbb{S} = n$), while $\xi Q = 0$ is a set of as many equations as unknowns, n, the detailed balance equations form a set of $n(n-1)/2$ different equations for n unknowns, so one would not expect solutions, in general. However, if the Q-matrix is sparse, i.e. contains lots of zeros, corresponding equations will be automatically satisfied, and these are the cases where we will successfully apply detailed balance equations.

The class of continuous-time Markov chains for which the detailed balance equations have solutions can be studied further. They also arise naturally in the context of time reversal, a tool that may seem of little practical relevance, since our world lives forward in time, but sometimes it is useful to model by a random process an unknown past. Sometimes, one can identify a duality relationships between two different processes, both forward in time that reveals that the behaviour of one is the same as the behaviour of the time reversal of the other. This can allow to translate known results for one into interesting new results for the other.

Proposition 57 *Let X be an irreducible positive recurrent (minimal) continuous-time Markov chain with Q-matrix Q and starting from the invariant distribution ξ. Let $t > 0$ be a fixed time and $\hat{X}_s = X_{t-s-}$. Then the process $\hat{X}$ is a continuous-time Markov chain with Q-matrix $\hat{Q}$ given by $\xi_j \hat{q}_{ji} = \xi_i q_{ij}$.*

Proof: First note that $\hat{Q}$ has the properties of a Q-matrix in being non-negative off the diagonal and satisfying

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

by the invariance of ξ. Similarly, we define $\xi_j \hat{p}_{ji}(t) = \xi_i p_{ij}(t)$ and see that $\hat{P}(t)$ have the properties of transition matrices. In fact the transposed forward equation $P'(t) = P(t)Q$ yields $\hat{P}'(t) = \hat{Q}\hat{P}(t)$, the backward equation for $\hat{P}(t)$. Now $\hat{X}$ is a continuous-time Markov chain with transition probabilities $\hat{P}(t)$ since

$$\begin{aligned} \mathbb{P}_\xi(\hat{X}_{t_0} = i_0, \ldots, \hat{X}_{t_n} = i_n) &= \mathbb{P}_\xi(X_{t-t_n} = i_n, \ldots, X_{t-t_0} = i_0) \\ &= \xi_{i_n} \prod_{k=1}^{n} p_{i_k, i_{k-1}}(t_k - t_{k-1}) \\ &= \xi_{i_0} \prod_{k=1}^{n} \hat{p}_{i_{k-1}, i_k}(t_k - t_{k-1}). \end{aligned}$$

From this we can deduce the Markov property. More importantly, the finite-dimensional distributions of $\hat{X}$ are the ones of a continuous-time Markov chain with transition matrices $\hat{P}(t)$. Together with the path structure, Remark 36 implies that $\hat{X}$ is a Markov chain with Q-matrix $\hat{Q}$. □

If $\hat{Q} = Q$, X is called *reversible.* It is evident from the definition of $\hat{Q}$ that ξ then satisfies the *detailed balance equations* $\xi_i q_{ij} = \xi_j q_{ji}$.

8.4 Erlang's formula

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

We recognise birth processes and queueing systems as special cases.

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

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

or more easily the detailed balance equations

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

giving

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

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

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

provided S is finite.

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

On the other hand, if $\lambda_0 = 0$ as in many population models, then the invariant distribution is concentrated in 0, i.e. $\xi_0 = 1$, $\xi_n = 0$ for all $n \geq 1$.

Many special cases can be given more explicitly. If $\lambda_n = \lambda$, $n \geq 0$, $\mu_n = n\mu$, we get

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

You recognise the Poisson probabilities.