Chapter 2

Continuous time Markov chains

As before we assume that we have a finite or countable statespace $I$, but now the Markov chains $X = \{X(t) : t \geq 0\}$ have a continuous time parameter $t \in [0, \infty)$. In some cases, but not the ones of interest to us, this may lead to analytical problems, which we skip in this lecture.

2.1 Q-Matrices

In continuous time there are no smallest time steps and hence we cannot speak about one-step transition matrices any more. If we are in state $j \in I$ at time $t$, then we can ask for the probability of being in a different state $k \in I$ at time $t + h$,

$$f(h) = \mathbb{P}\{X(t + h) = k | X(t) = j\}.$$  

We are interested in small time steps, i.e. small values of $h > 0$. Clearly, $f(0) = 0$. Assuming that $f$ is differentiable at 0, the most interesting information we can obtain about $f$ at the origin is its derivative $f'(0)$, which we call $q_{jk}$.

$$\lim_{h \downarrow 0} \frac{\mathbb{P}\{X(t + h) = k | X(t) = j\}}{h} = \lim_{h \downarrow 0} \frac{f(h) - f(0)}{h - 0} = f'(0) = q_{jk}.$$  

We can write this as

$$\mathbb{P}\{X(t + h) = k | X(t) = j\} = q_{jk}h + o(h).$$  

Here $o(h)$ is a convenient abbreviation for any function with the property that $\lim_{h \downarrow 0} o(h)/h = 0$, we say that the function is of smaller order than $h$. An advantage of this notation is that the actual function $o$ might be a different one in each line, but we do not need to invent a new name for each occurrence.

Again we have the Markov property, which states that if we know the state $X(t)$ then all additional information about $X$ at times prior to $t$ is irrelevant for the future: For all $k \neq j$, $t_0 < t_1 < \ldots < t_n < t$ and $x_0, \ldots, x_n \in I$,

$$\mathbb{P}\{X(t + h) = k | X(t) = j, X(t_i) = x_i \forall i\} = q_{jk}.$$  

1
From this we get, for every $j \in I$,
\[
P\{X(t+h) = j \mid X(t) = j\} = 1 - \sum_{k \in I} q_{jk} h + o(h) = 1 + q_{jj} h + o(h),
\]
if we define $q_{jj} = -\sum_{k \in I} q_{jk}$. As above we have $q_{jj} = f'(0)$ for $f(h) = P\{X(t+h) = j \mid X(t) = j\}$, but now $f(0) = 1$.

We can enter this information into a matrix $Q = (q_{ij} : i, j \in I)$, which contains all the information about the transitions of the Markov chain $X$. This matrix is called the $Q$-matrix of the Markov chain. Its necessary properties are

- all off-diagonal entries $q_{ij}$, $i \neq j$, are non-negative,
- all diagonal entries $q_{ii}$ are nonpositive and,
- the sum over the entries in each row is zero (!).

We say that $q_{ij}$ gives the rate at which we try to enter state $j$ when we are in state $i$, or the jump intensity from $i$ to $j$.

**Example 2.1: The Poisson process**

Certain random occurrences (for example claim times to an insurance company, arrival times of customers in a shop, . . . ) happen randomly in such a way that

- the probability of one occurrence in the time interval $(t, t + h)$ is $\lambda h + o(h)$, the probability of more than one occurrence is $o(h)$.
- occurrences in the time interval $(t, t + h)$ are independent of what happened before time $t$.

Suppose now that $X(t)$ is the number of occurrences by time $t$. Then $X$ is a continuous time Markov chain with statespace $I = \{0, 1, 2, \ldots\}$. Because, by our assumptions,
\[
P\{X(t+h) = k \mid X(t) = j\} = \begin{cases} \lambda h + o(h) & \text{if } k = j + 1, \\ 0 + o(h) & \text{if } k > j + 1, \\ 0 & \text{if } k < j, \end{cases}
\]
we get the $Q$-matrix
\[
Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.
\]
This process is called a **Poisson process with rate** $\lambda$.

Reason for this name: An analysis argument shows that, if $X(0) = 0$, then
\[
P\{X(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \text{ for } n = 0, 1, 2, \ldots,
\]
i.e. $X(t)$ is Poisson distributed with parameter $\lambda t$.

**Example 2.2: The pure birth-process**

Suppose we have a population of (immortal) animals reproducing in such a way that, independent of what happened before time $t$ and of what happens to the other animals, in the interval $(t, t + h)$ each animal
• gives birth to 1 child with probability $\lambda h + o(h)$,
• gives birth to more than 1 child with probability $o(h)$,
• gives birth to 0 children with probability $1 - \lambda h + o(h)$.

Let $p = \lambda h + o(h)$ and $q = 1 - p$ and $X(t)$ the number of animals at time $t$. Then,

$$\mathbb{P}\{X(t+h) = n \mid X(t) = n\} = \mathbb{P}\{\text{no births in } (t, t+h) \mid X(t) = n\}$$

$$= q^n = [1 - \lambda h + o(h)]^n$$

$$= 1 - n\lambda h + o(h),$$

and

$$\mathbb{P}\{X(t+h) = n + 1 \mid X(t) = n\} = \mathbb{P}\{\text{one birth in } (t, t+h) \mid X(t) = n\}$$

$$= npq^{n-1} = n\lambda h [1 - \lambda h + o(h)]^{n-1}$$

$$= n\lambda h + o(h),$$

and finally, for $k > 1$,

$$\mathbb{P}\{X(t+h) = n + k \mid X(t) = n\} = \mathbb{P}\{k \text{ births in } (t, t+h) \mid X(t) = n\}$$

$$= \binom{n}{k} p^k q^{n-k} + o(h) = \binom{n}{k} \lambda^k h^k [1 - \lambda h + o(h)]^{n-k} + o(h)$$

$$= o(h).$$

Hence, $X$ is a continuous time Markov chain with $Q$-matrix

$$Q = \begin{pmatrix}
-\lambda & \lambda & 0 & 0 & \ldots \\
0 & -2\lambda & 2\lambda & 0 & \ldots \\
0 & 0 & -3\lambda & 3\lambda & 0 & \ldots \\
0 & \ldots & 0 & \ddots & \ddots & \ddots \\
\end{pmatrix}.$$

An analysis argument shows that, if $X(0) = 1$, then

$$\mathbb{P}\{X(t) = n\} = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1} \text{ for } n = 1, 2, 3, \ldots,$$

i.e. $X(t)$ is geometrically distributed with parameter $e^{-\lambda t}$.

### 2.1.1 Construction of the Markov chain

Given the $Q$-matrix one can construct the paths of a continuous time Markov chain as follows. Suppose the chain starts in a fixed state $X_0 = i$ for $i \in I$. Let

$$J = \min\{t : X_t \neq X_0\}$$

be the first jump time of $X$ (we always assume that the minimum exists).
Theorem 2.1 Under the law $P_i$ of the Markov chain started in $X_0 = i$ the random variables $J$ and $X(J)$ are independent. The distribution of $J$ is exponential with rate $q_i := \sum_{j \neq i} q_{ij}$, which means

$$P_i\{J > t\} = e^{-q_i t} \text{ for } t \geq 0.$$ 

Moreover,

$$P_i\{X(J) = j\} = \frac{q_{ij}}{q_i},$$

and the chain starts afresh at time $J$.

### Picture!

Let $J_1, J_2, J_3, \ldots$ be the times between successive jumps. Then

$$Y_n = X(J_1 + \cdots + J_n)$$

defines a discrete time Markov chain with one-step transition matrix $P$ given by

$$p_{ij} = \begin{cases} q_{ij} / q_i & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Given $Y_n = i$ the next waiting time $J_{n+1}$ is exponentially distributed with rate $q_i$ and independent of $Y_1, \ldots, Y_{n-1}$ and $J_1, \ldots, J_n$.

### 2.1.2 Exponential times

Why does the exponential distribution play a special role for continuous Markov chains?

Recall the Markov property in the following way: Suppose $X(0) = i$ and let $J$ be the time before the first jump and $t, h > 0$. Then, using the Markov property in the second step,

$$P\{J > t + h \mid J > t\} = P\{J > t + h \mid J > t, X(t) = i\} = P\{J > t + h \mid X(t) = i\} = P\{J > h\}.$$ 

Hence the time $J$ we have to wait for a jump satisfies the lack of memory property: if you have waited for $t$ time units and no jump has occurred, the remaining waiting time has the same distribution as the original waiting time. This is sometimes called the waiting time paradox or constant failure rate property.

The only distribution with the lack of memory property is the exponential distribution. Check that, for $P\{J > x\} = e^{-\mu x}$, we have

$$P\{J > t + h \mid J > t\} = \frac{e^{-\mu(t+h)}}{e^{-\mu t}} = e^{-\mu h} = P\{J > h\}.$$

Another important property of the exponential distribution is the following:

**Theorem 2.2** If $S$ and $T$ are independent exponentially distributed random variables with rate $\alpha$ resp. $\beta$, then their minimum $S \wedge T$ is also exponentially distributed with rate $\alpha + \beta$ and it is independent of the event $\{S \wedge T = S\}$. Moreover,

$$P\{S \wedge T = S\} = \frac{\alpha}{\alpha + \beta} \text{ and } P\{S \wedge T = T\} = \frac{\beta}{\alpha + \beta}.$$

Prove this as an exercise, see Sheet 9!
Example 2.3: The M/M/1 queue.

Suppose customers arrive according to a Poisson process with of rate $\lambda$ at a single server. Each customer requires an independent random service time, which is exponential with mean $1/\mu$ (i.e. with rate $\mu$). Let $X(t)$ be the number of people in the queue (including people currently being served). Then $X$ is a continuous-time Markov chain with

$$Q = \begin{pmatrix}
-\lambda & \lambda & 0 & 0 & 0 & \ldots \\
0 & -\lambda - \mu & \lambda & 0 & 0 & \ldots \\
0 & \mu & -(\lambda + \mu) & \lambda & 0 & \ldots \\
0 & 0 & \ldots & \ddots & \ddots & \ddots & \ldots \\
\end{pmatrix}.$$

The previous theorem says: If there is at least one person in the queue, the distribution of the jump time (i.e. the time until the queue changes) is exponential with rate $\lambda + \mu$ and the probability that at the jump time the queue is getting shorter is $\mu/(\lambda + \mu)$.

2.2 Kolmogorov’s equations and global Markov property

Apart from the local version of the Markov property, for small $h > 0$,

$$P(h) = \{ p_{ij}(t), i, j \in I, t \geq 0 \} \text{ given by } p_{ij}(t) = P_i\{ X(t) = j \},$$

continuous-time Markov chains also satisfy a global Markov property. To describe it let

$$P(t) = \{ p_{ij}(t), i, j \in I, t \geq 0 \} \text{ given by } p_{ij}(t) = P_i\{ X(t) = j \},$$

be the transition matrix function of the Markov chain. From $P$ one can get the full information about the law of the Markov chain, for $0 < t_1 < \ldots < t_n$ and $j_1, \ldots, j_n \in I$,

$$P_i\{ X(t_1) = j_1, \ldots, X(t_n) = j_n \} = p_{i_1j_1}(t_1)p_{j_1j_2}(t_2 - t_1)\cdots p_{j_{n-1}j_n}(t_n - t_{n-1}).$$

$P$ has the following properties,

(a) $p_{ij}(t) \geq 0$ and $\sum_{j \in I} p_{ij}(t) = 1$ for all $i \in I$. Also

$$\lim_{t \to 0} p_{ij}(t) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

(b) $\sum_{k \in I} p_{ik}(t)p_{kj}(s) = p_{ij}(t + s)$ for all $i, j \in I$ and $s, t \geq 0$.

The equation in (b) is called the Chapman-Kolmogorov equation. It is easily proved,

$$P_i\{ X(t + s) = j \} = \sum_{k \in I} P_i\{ X(t) = k, X(t + s) = j \}$$

$$= \sum_{k \in I} P_i\{ X(t) = k \} P_i\{ X(t + s) = j | X(t) = k \}$$

$$= \sum_{k \in I} P_i\{ X(t) = k \} P_k\{ X(s) = j \}.$$
Note that, by definition $p_{jk}(h) = q_{jk}h + o(h)$ for $j \neq k$, and hence

$$p'_{jk}(0) = q_{jk} \text{ and } p'_{jj}(0) = q_{jj}. $$

From this we can see how the $Q$-matrix is obtainable from the transition matrix function. The converse operation is more involved, we restrict attention to the case of finite statespace $I$.

Consider $P(t)$ as a matrix, then the Chapman-Kolmogorov equation can be written as

$$P(t + s) = P(t)P(s) \text{ for all } s, t \geq 0.$$

One can differentiate this equation with respect to $s$ or $t$ and gets

$$P'(t) = QP(t) \text{ and } P'(t) = P(t)Q.$$ 

These equations are called the Kolmogorov backward resp. Kolmogorov forward equations. We also know $P(0) = I$, the identity matrix. This matrix-valued differential equation has a unique solution $P(t), t \geq 0$.

For the case of a scalar function $p(t)$ and a scalar $q$ instead of the matrix-valued function $P$ and matrix $Q$ the corresponding equation

$$p'(t) = qp(t), \ p(0) = 1,$$

would have the unique solution $p(t) = e^{qt}$. In the matrix case one can argue similarly, defining

$$e^{Qt} = \sum_{k=0}^{\infty} \frac{Q^k t^k}{k!},$$

with $Q^0 = I$. Then we get the transition matrix function $P(t) = e^{Qt}$ which satisfies the Kolmogorov forward and backward equations.

### 2.3 Resolvents

Let $Q$ be the $Q$-matrix of the Markov chain $X$ and $P(t), t \geq 0$ the transition matrix function.

#### 2.3.1 Laplace transforms

A function $f : [0, \infty) \to \mathbb{R}$ is called good if $f$ is continuous and either bounded (there exists $K > 0$ such that $|f(t)| \leq K$ for all $t$) or integrable ($\int_0^\infty |f(t)| \, dt < \infty$) or both. If $f$ is good we can associate the Laplace transform $\hat{f} : (0, \infty) \to \mathbb{R}$ which is defined by

$$\hat{f}(\lambda) = \int_0^{\infty} e^{-\lambda t} f(t) \, dt.$$ 

An important example is the case $f(t) = e^{-\alpha t}$. Then

$$\hat{f}(\lambda) = \int_0^{\infty} e^{-\lambda t} e^{-\alpha t} \, dt = \frac{1}{\lambda + \alpha}.$$ 

An important result about Laplace transforms is the following:
Theorem 2.3 (Uniqueness Theorem) If \( f \) and \( g \) are good functions on \([0, \infty)\) and \( \hat{f}(\lambda) = \hat{g}(\lambda) \) for all \( \lambda > 0 \). Then \( f = g \).

This implies that we can invert Laplace transforms uniquely, at least when restricting attention to good functions.

2.3.2 Resolvents

The basic idea here is to calculate the exponential \( P(t) = e^{Qt} \) of the \( Q \)-matrix using the Laplace transform. Let \( \lambda > 0 \) and argue that

\[
R(\lambda) := \int_0^\infty e^{-\lambda t} P(t) dt = \int_0^\infty e^{-\lambda t} e^{Qt} dt = \int_0^\infty e^{(\lambda I - Q)t} dt = (\lambda I - Q)^{-1}.
\]

Here the integrals over matrix-valued functions are understood componentwise, and the last step is done by analogy with the real-valued situation, but can be made rigorous. Note that the inverse of the matrix \( \lambda I - Q \) can be calculated for all values \( \lambda \) which are \textit{not an eigenvalue of} \( Q \). Recall that the inverse fails to exist if and only if the characteristic polynomial \( \det(\lambda I - Q) \) is zero, which is also the criterion for \( \lambda \) to be an eigenvalue of \( Q \).

Now \( R(\lambda) = (\lambda I - Q)^{-1} \), if it exists, can be calculated and \( P(t) \) can be recovered by inverting the Laplace transform. The matrix function \( R(\lambda) \) is called the \textit{resolvent} of \( Q \) (or of \( X \)). Its components are the Laplace transforms of the functions \( p_{ij} \),

\[
\boxed{r_{ij}(\lambda) := \int_0^\infty e^{-\lambda t} p_{ij}(t) dt = \hat{p}_{ij}(\lambda).}
\]

Resolvents also have a probabilistic interpretation: Suppose we have an alarm-clock which rings, independently of the Markov chain, at a random time \( A \), which is exponentially distributed with rate \( \lambda \), i.e.

\[
P\{A > t\} = e^{-\lambda t}, \quad P\{A \in (t, t + dt)\} = \lambda e^{-\lambda t} dt.
\]

What is the state of the Markov chain when the alarm clock rings? The probability of being in state \( j \) is

\[
P_i\{X(A) = j\} = \int_0^\infty P_i\{X(t) = j, A \in (t, t + dt)\} = \int_0^\infty p_{ij}(t)\lambda e^{-\lambda t} dt = \lambda r_{ij}(\lambda),
\]

hence

\[
P_i\{X(A) = j\} = \lambda r_{ij}(\lambda) \text{ for } B \text{ independent of } X, \text{ exponential with rate } \lambda.
\]

We now discuss the use of resolvents in calculations for continuous-time Markov chains using the following problem:

Problem: Consider a continuous-time Markov chain \( X \) with state space \( I := \{A, B, C\} \) and transitions between the states described as follows:

- When currently in state \( A \) at time \( t \), in the next time interval \((t, t+h)\) of \textit{small} length \( h \) and independent of the past behaviour of the chain, with
probability $h + o(h)$ the chain jumps into state $B$,
probability $2h + o(h)$ the chain jumps into state $C$,
probability $1 - 3h + o(h)$ the chain remains in state $A$.

- Whilst in state $B$, the chain tries to enter state $C$ at rate 2. The chain cannot jump into state $A$ directly from state $B$.

- On entering state $C$, the chain remains there for an independent exponential amount of time of rate 3 before jumping. When the jump occurs, with probability $2/3$ it is into state $A$, and with probability $1/3$ the jump is to state $B$.

Questions:

(a) Give the $Q$-matrix of the continuous-time Markov chain $X$.

(b) If the Markov chain is initially in state $A$, that is $X(0) = A$, what is the probability that the Markov chain is still in state $A$ at time $t$, that is, $P_A\{X(t) = A\}$?

(c) Starting from state $C$ at time 0, show that the probability $X$ is in state $B$ at time $t$ is given by

$$\frac{1}{3} - \frac{1}{3}e^{-3t} \text{ for } t \geq 0.$$ 

(d) Initially, $X$ is in state $C$. Find the distribution for the position of $X$ after an independent exponential time, $T$, of rate 4 has elapsed. In particular, you should find

$$P_C\{X(T) = B\} = \frac{1}{7}.$$ 

(e) Given that $X$ starts in state $C$, find the probability that $X$ enters state $A$ at some point before time $t$. [See Section 2.3.3 for solution.]

(f) What is the probability that we have reached state $C$ at some point prior to time $U$, where $U$ is an independent exponential random variable of rate $\mu = 2.5$, given that the chain is in state $B$ at time zero? [See Section 2.3.3 for solution.]

The method to obtain (a) should be familiar to you by now. For the jumps from state $C$ we use Theorem 2.2. Recall $q_{CA}$ is the rate of jumping from $C$ to $A$ and $q_{CB}$ is the rate of jumping from $C$ to $B$. We are given $3 = q_{CA} + q_{CB}$, the rate of leaving state $C$. The probability that the first jump goes to state $A$ is, again by Theorem 2.2, $q_{CA}/(q_{CA} + q_{CB})$, which is given as $2/3$. Hence we get $q_{CA} = 2$ and $q_{CB} = 1$. Altogether,

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

8
We now solve (b) with the resolvent method. Recall
\[ p_{AA}(t) = P_A \{ X(t) = A \}, \text{ and } r_{AA}(t) = \int_0^\infty e^{-\lambda t} p_{AA}(t) \, dt = \hat{p}_{AA}(t). \]

As \( R(\lambda) = (\lambda I - Q)^{-1} \) we start by inverting
\[ \lambda I - Q = \begin{pmatrix} \lambda + 3 & -1 & -2 \\ 0 & \lambda + 2 & -2 \\ -2 & -1 & \lambda + 3 \end{pmatrix}. \]

First,
\[ \det(\lambda I - Q) = (\lambda + 2) \det \left( \begin{array}{cc} \lambda + 3 & -2 \\ -2 & \lambda + 3 \end{array} \right) + 2 \det \left( \begin{array}{cc} \lambda + 3 & -1 \\ -2 & -1 \end{array} \right) = \lambda(\lambda + 3)(\lambda + 5). \]

(Side remark: \( \lambda \) is always a factor in this determinant, because \( Q \) is a singular matrix. Recall that the sum over the rows is zero!) Now the inverse of a matrix \( A \) is given by
\[ (A^{-1})_{ij} = \frac{(-1)^{i+j}}{\det A} \text{det}(M_{ji}), \]
where \( M_{ji} \) is the matrix with row \( j \) and column \( i \) removed. Hence, in our example, the entry in position \( AA \) is obtained by
\[ r_{AA}(\lambda) = (\lambda I - Q)^{-1}_{AA} = \frac{\det \left( \begin{array}{cc} \lambda + 2 & -2 \\ -1 & \lambda + 3 \end{array} \right)}{\det(\lambda I - Q)} = \frac{\lambda^2 + 5\lambda + 4}{\lambda(\lambda + 3)(\lambda + 5)}. \]

It remains to invert the Laplace transform. For this purpose we need to form partial fractions. Solving
\[ \lambda^2 + 5\lambda + 4 = \alpha(\lambda + 3)(\lambda + 5) + \beta\lambda(\lambda + 5) + \gamma\lambda(\lambda + 3) \]
gives \( \alpha = 4/15 \) (plug \( \lambda = 0 \)) and \( \beta = 1/3 \) (plug \( \lambda = -3 \)) and \( \gamma = 2/5 \) (plug \( \lambda = -5 \)). Hence
\[ r_{AA}(\lambda) = \frac{4}{15\lambda} + \frac{1}{3(\lambda + 3)} + \frac{2}{5(\lambda + 5)}. \]

Now the inverse Laplace transform of \( 1/(\lambda + \beta) \) is \( e^{-\beta t} \). We thus get
\[ p_{AA}(t) = \frac{4}{15} + \frac{1}{3} e^{-3t} + \frac{2}{15} e^{-5t}, \]
and note that \( p_{AA}(0) = 1 \) as it should be!

Now solve part (c). To find \( P_B \{ X(t) = B \} \) consider
\[ r_{CB}(\lambda) = \int_0^\infty e^{-\lambda t} p_{CB}(t) \, dt = \hat{p}_{CB}(t). \]

We need to find the \( CB \) coefficient of the matrix \( (\lambda I - Q)^{-1} \), recalling the inversion formula we get
\[ r_{CB}(\lambda) = \frac{-\det \left( \begin{array}{cc} \lambda + 3 & -1 \\ -2 & -1 \end{array} \right)}{\lambda(\lambda + 3)(\lambda + 5)} = \frac{\lambda + 5}{\lambda(\lambda + 3)(\lambda + 5)} = \frac{1}{\lambda(\lambda + 3)} = \frac{1}{3} \left( \frac{1}{\lambda} - \frac{1}{\lambda + 3} \right), \]

\( 9 \)
and inverting Laplace transforms gives

\[ \mathbf{P}_C \{ X(t) = B \} = p_{CB}(t) = \frac{1}{3} - \frac{1}{3}e^{-3t}. \]

Suppose for part (d) that \( T \) is exponentially distributed with rate \( \lambda \) and independent of the chain. Recall,

\[ \mathbf{P}_i \{ X(T) = j \} = \lambda r_{ij}(\lambda), \]

and by matrix inversion,

\[ r_{CA}(\lambda) = \frac{\det \begin{pmatrix} 0 & \lambda + 2 \\ -2 & -1 \end{pmatrix}}{\det(\lambda I - Q)} = \frac{2(\lambda + 2)}{\lambda(\lambda + 3)(\lambda + 5)}. \]

As \( T \) is exponential with parameter \( \lambda = 4 \) we get (no Laplace inversion needed!)

\[ \mathbf{P}_C \{ X(T) = A \} = \frac{2(4 + 2)}{(4 + 3)(4 + 5)} = \frac{12}{63} = \frac{4}{21}. \]

Similarly,

\[ r_{CB}(\lambda) = \frac{-\det \begin{pmatrix} \lambda + 3 & -1 \\ -2 & -1 \end{pmatrix}}{\det(\lambda I - Q)} = \frac{\lambda + 5}{\lambda(\lambda + 3)(\lambda + 5)}, \]

and

\[ r_{CC}(\lambda) = \frac{\det \begin{pmatrix} \lambda + 3 & -1 \\ 0 & \lambda + 2 \end{pmatrix}}{\det(\lambda I - Q)} = \frac{(\lambda + 3)(\lambda + 2)}{\lambda(\lambda + 3)(\lambda + 5)}. \]

Plugging \( \lambda = 4 \) we get

\[ \mathbf{P}_C \{ X(T) = B \} = 4r_{CB}(4) = \frac{1}{7} \quad \text{and} \quad \mathbf{P}_C \{ X(T) = C \} = \frac{4 + 2}{4 + 5} = \frac{2}{3}. \]

At this point it is good to check that the three values add up to one and thus define a probability distribution on the statespace \( I = \{ A, B, C \} \).

### 2.3.3 First hitting times

Let \( T_j = \inf\{ t > 0 : X_t = j \} \) be the first positive time we are in state \( j \). Define, for \( i, j \in I \),

\[ F_{ij}(t) = \mathbf{P}_i \{ T_j \leq t \} = \mathbf{P}\{ T_j \leq t \mid X(0) = i \}, \]

then \( F_{ij}(t) \) is the probability that we have entered state \( j \) at some time prior to time \( t \), if we start the chain in \( i \). We have the first hitting time matrix function,

\[ F(t) = \left( F_{ij}(t) : i, j \in I, t \geq 0 \right), \]

with \( F(0) = I \) and

\[ F_{ij}(t) = \int_0^t f_{ij}(s) \, ds = \int_0^t \mathbf{P}_i \{ T_j \in ds \}, \]

10
where \( f_{ij}(t) \) is the first hitting time density. The matrix functions \( f = (f_{ij}(t), t \geq 0) \) and \( P \) are linked by the integral equation

\[
p_{ij}(t) = \int_0^t f_{ij}(s)p_{jj}(t-s)\,ds, \quad \text{for } i, j \in I \text{ and } t \geq 0. \tag{2.3.1}
\]

This can be checked as follows, use the strong Markov property, which is the fact that the chain starts afresh at time \( T_j \) in state \( j \) and evolves independently of everything that happened before time \( T_j \) to get

\[
p_{ij}(t) = P_i\{X_t = j\} = \int_0^t P_i\{X_t = j, T_j \leq t\} = \int_0^t f_{ij}(s)p_{jj}(t-s)\,ds.
\]

Given two functions \( f, g : [0, \infty) \to \mathbb{R} \) we can define the convolution function \( f \ast g \) by

\[
f \ast g(t) = \int_0^t f(s)g(t-s)\,ds,
\]

and check by substituting \( u = t-s \) that \( f \ast g = g \ast f \). We can rewrite the integral equation (2.3.1) as

\[
p_{ij}(t) = f_{ij} \ast p_{jj}(t).
\]

In order to solve this equation for the unknown \( f_{ij} \) we use Laplace transforms again.

**Theorem 2.4 (Convolution Theorem)** If \( f, g \) are good functions, then \( \hat{f \ast g}(\lambda) = \hat{f}(\lambda)\hat{g}(\lambda) \).

To make this plausible look at the example of two functions \( f(t) = e^{-\alpha t} \) and \( g(t) = e^{-\beta t} \). Then

\[
f \ast g(t) = \int_0^t f(s)g(t-s)\,ds = e^{-\beta t} \int_0^t e^{-(\alpha-\beta)s}\,ds = \frac{e^{-\beta t} - e^{-\alpha t}}{\alpha - \beta}.
\]

Then

\[
\hat{f \ast g}(\lambda) = \int_0^\infty e^{-\lambda s}(f \ast g)(s)\,ds = \frac{1}{\alpha - \beta} \left\{ \frac{1}{\lambda + \beta} - \frac{1}{\lambda + \alpha} \right\}
\]

\[
= \frac{1}{\lambda + \alpha} \frac{1}{\lambda + \beta} = \hat{f}(\lambda)\hat{g}(\lambda).
\]

The convolution theorem applied to the integral equation (2.3.1) gives a formula for the Laplace transform of the first hitting time density

\[
\hat{f}_{ij}(\lambda) = \frac{r_{ij}(\lambda)}{r_{jj}(\lambda)} \quad \text{for } i, j \in I \text{ and } \lambda > 0. \tag{2.3.2}
\]

We use this now to given an answer to (e) in our problem. We are looking for

\[
P_C\{T_A \leq t\} = F_{CA}(t) = \int_0^t f_{CA}(s)\,ds.
\]
We first find the Laplace transform $\hat{f}_{CA}$ of $f_{CA}$, using $r_{CA}$ and $r_{AA}$,

$$
\hat{f}_{CA} = \frac{\det \begin{pmatrix} 0 & \lambda + 2 \\ -2 & -1 \end{pmatrix}}{\det \begin{pmatrix} \lambda + 2 & -2 \\ -1 & \lambda + 3 \end{pmatrix}} = \frac{2(\lambda + 2)}{\lambda^2 + 5\lambda + 4},
$$

and by partial fractions

$$
\hat{f}_{CA} = \frac{2}{3(\lambda + 1)} + \frac{4}{3(\lambda + 4)}.
$$

In this form, the Laplace transform can be inverted, which gives

$$
f_{CA}(t) = \frac{2}{3}e^{-t} + \frac{4}{3}e^{-4t} \text{ for } t \geq 0.
$$

Now, by integration, we get

$$
P_C\{T_A \leq t\} = \int_0^t f_{CA}(s) \, ds = \frac{2}{3} \int_0^t e^{-s} \, ds + \frac{4}{3} \int_0^t e^{-4s} \, ds = 1 - \frac{2}{3}e^{-t} - \frac{1}{3}e^{-4t}.
$$

Check that $P_C\{T_A \leq 0\} = 0$, as expected.

As in the case of $r_{ij} = \hat{p}_{ij}$ the Laplace transforms $\hat{f}_{ij}$ also admit a direct probabilistic interpretation with the help of an alarm clock, which goes off at a random time $A$, which is independent of the Markov chain $X$ and exponentially distributed with rate $\lambda$. Indeed, recalling that $\lambda e^{-x\lambda}, x > 0$ is the density of $A$ and $f_{ij}$ is the density of $T_j$,

$$
P_i\{T_j \leq a\} = \int_0^\infty P_i\{T_j \leq a\} \lambda e^{-a\lambda} \, da
= \int_0^\infty \int_0^a f_{ij}(s) \, ds \lambda e^{-a\lambda} \, da
= \int_0^\infty f_{ij}(s) \int_s^\infty \lambda e^{-a\lambda} \, da \, ds
= \int_0^\infty f_{ij}(s) e^{-a\lambda} \, da
= \hat{f}_{ij}(\lambda).
$$

Altogether,

$$
\boxed{\hat{f}_{ij}(\lambda) = P_i\{T_j < A\} \text{ for } i, j \in I \text{ and } \lambda > 0.}
$$

We use this now to given an answer to (f) in our problem. The question asks for $P_B\{T_C < U\}$ where $U$ is exponential with rate $\mu = 2.5$ and independent of the Markov chain. We now know that

$$
P_B\{T_C < U\} = \hat{f}_{BC}(\mu) = \frac{r_{BC}(\mu)}{r_{CC}(\mu)} = \frac{-\det \begin{pmatrix} \mu + 1 & -2 \\ 0 & -2 \end{pmatrix}}{\det \begin{pmatrix} \mu + 1 & -1 \\ 0 & \mu + 2 \end{pmatrix}} = \frac{2(\mu + 3)}{(\mu + 3)(\mu + 2)} = \frac{2}{\mu + 2}.
$$
and plugging in $\mu = 2.5$ gives $P_B\{T_C < U\} = \frac{4}{\pi}$.

**Example:** A machine can be in three states $A, B, C$ and the transition between the states is given by a continuous-time Markov chain with $Q$-matrix

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

At time zero the machine is in state $A$. Suppose a supervisor arrives for inspection at the site of the machine at an independent exponential random time $T$, with rate $\lambda = 2$. What is the probability that he has arrived when the machine is in state $C$ for the first time?

We write $T_C = \inf\{t > 0 : X(t) = C\}$ for the first entry and $S_C = \inf\{t > T_C : X(t) \neq C\}$ for the first exit time from $C$. Then the required probability is $P_A\{T_C < T, S_C > T\}$. Note that $J = S_C - T_C$ is the length of time spent by the chain in state $C$ during the first visit. This time is exponentially distributed with rate $q_{CA} + q_{CB} = 3$ and we get

$$P_A\{T_C < T, S_C > T\} = P_A\{T_C < T\}P_A\{S_C > T | T_C < T\} = P_A\{T_C < T\}P_A\{T - T_C < J | T > T_C\}.$$ 

By the lack of memory property, given $T > T_C$ the law of $\tilde{T} = T - T_C$ is exponential with rate $\lambda = 2$ again. Hence

$$P_A\{T_C < T, S_C > T\} = P_A\{T_C < T\}P_C\{\tilde{T} < J\},$$ 

where $\tilde{T}$ is an independent exponential random variable with rate $\lambda = 2$. The right hand side can be calculated. We write down

$$\lambda I - Q = \begin{pmatrix} \lambda + 3 & -1 & -2 \\ -3 & \lambda + 3 & 0 \\ -1 & -2 & \lambda + 3 \end{pmatrix},$$

and derive

$$P_A\{T_C < T\} = \frac{\lambda}{\lambda + 3} = \frac{\det \begin{pmatrix} -1 & -2 \\ \lambda + 3 & 0 \end{pmatrix}}{\det \begin{pmatrix} \lambda + 3 & -1 \\ -3 & \lambda + 3 \end{pmatrix}} = \frac{2(\lambda + 3)}{\lambda^2 + 6\lambda + 6},$$

and, by Theorem 2.2,

$$P_C\{J > \tilde{T}\} = \frac{\lambda}{\lambda + 3}$$

recalling that $J$ and $\tilde{T}$ are independent exponentially distributed with rates 3 resp. $\lambda = 2$. Hence,

$$P_A\{T_C < T, S_C > T\} = \frac{2\lambda}{\lambda^2 + 6\lambda + 6} = \frac{4}{22} = \frac{2}{11}.$$ 

### 2.4 Long-term behaviour and invariant distribution

We now look at the long-term behaviour of *continuous* time Markov chains. The notions of an *irreducible* and *recurrent* Markov chain can be defined in the same manner as in the discrete case:
A chain is irreducible if one can get from any state to any other state in finite time, and

an irreducible chain is recurrent if the probability that we return to this state in finite time is one.

We assume in this chapter that our chain satisfies these two conditions and also a technical condition called non-explosive, which ensures that the process is defined at all times, see Norris for further details.

2.4.1 The invariant distribution

Recall that in the discrete time case the stationary distribution \( \pi \) on \( I \) was given as the solution of the equation \( \pi P = \pi \) and, by iteration,

\[
\pi = \pi P = \pi P^2 = \pi P^3 = \cdots = \pi P^n \quad \text{for all } n \geq 0.
\]

Hence we would expect an invariant distribution \( \pi \) of a continuous time Markov chain with transition matrix function \( (P(t) : t \geq 0) \) to satisfy

\[
\pi P(t) = \pi \quad \text{for all } t \geq 0.
\]

If the statespace is finite, we can differentiate with respect to time to get \( \pi P'(t) = 0 \). Setting \( t = 0 \) and recalling \( P'(0) = Q \) we get

\[
\pi Q = 0.
\]

The two boxed statements can be shown to be equivalent in the general case, see Norris Theorem 3.5.5. Any probability distribution \( \pi \) satisfying \( \pi Q = 0 \) is called an invariant distribution, or sometimes equilibrium or stationary distribution.

**Theorem 2.5** Suppose \( (X(t) : t \geq 0) \) satisfies our assumptions and \( \pi \) solves

\[
\sum_{i \in I} \pi_i q_{ij} = 0, \pi_j > 0 \quad \text{for all } j \in I, \quad \text{and} \quad \sum_{i \in I} \pi_i = 1,
\]

then,

\[
\lim_{t \to \infty} p_{ij}(t) = \pi_j \quad \text{for all } i \in I,
\]

and if \( V_j(t) = \int_0^t 1_{\{X(s) = j\}} \, ds \) is the time spent in state \( j \) up to time \( t \), we have

\[
\lim_{t \to \infty} \frac{V_j(t)}{t} = \pi_j \quad \text{with probability one.}
\]

This theorem is analogous to (parts of) the Big Theorem in the discrete time case. Note that it also implies that the invariant
2.4.2 Symmetrisability

Suppose \( m = (m_i : i \in I) \) satisfies \( m_i > 0 \). We say that \( Q \) is \( m \)-symmetrisable if we can find \( m_i > 0 \) with

\[
m_i q_{ij} = m_j q_{ji} \quad \text{for all } i, j \in I
\]

These are the detailed balance equations.

**Theorem 2.6** If we find an \( m \) solving the detailed balance equations such that \( M = \sum_{i \in I} m_i < \infty \), then \( \pi_i = m_i/M \) defines the unique invariant distribution.

For the proof note that, for all \( j \in I \),

\[
(\pi Q)_j = \frac{1}{M} \sum_{i \in I} m_i q_{ij} = \frac{1}{M} \sum_{i \in I} m_j q_{ji} = 0.
\]

Note that, as in the discrete case, the matrix \( Q \) may not be symmetrisable, but the invariant distribution may still exist. Then the invariant distribution may be found using generating functions.

**Example:** The M/M/1 queue.

A single server has a service rate \( \mu \), customers arrive individually at a rate \( \lambda \). Let \( X(t) \) be the number of cutomers in the queue (including the customer currently served) and \( I = \{0, 1, 2, \ldots\} \). The \( Q \)-matrix is given by

\[
Q = \begin{pmatrix}
-\lambda & \lambda & 0 & \cdots \\
\mu & -\lambda - \mu & \lambda & 0 & \cdots \\
0 & \mu & -\lambda - \mu & \lambda & 0 & \cdots \\
\cdots & 0 & \cdots & \cdots & \cdots & \cdots & 0
\end{pmatrix}.
\]

We first try to find the invariant distribution using symmetrisability. For this purpose we have to solve the detailed balance equations \( m_i q_{ij} = m_j q_{ji} \), explicitly

\[
m_0 \lambda = m_1 \mu \Rightarrow m_1 = \frac{\lambda}{\mu} m_0,
\]

\[
m_1 \lambda = m_2 \mu \Rightarrow m_2 = \frac{\lambda}{\mu} m_1,
\]

\[
m_n \lambda = m_{n+1} \mu \Rightarrow m_{n+1} = \frac{\lambda}{\mu} m_n,
\]

for all \( n \geq 2 \). Hence denoting by \( \rho = \lambda/\mu \) the traffic intensity we have \( m_i = \rho^i m_0 \). If \( \rho < 1 \), then

\[
M = \sum_{i=0}^{\infty} m_i = m_0 \sum_{i=0}^{\infty} \rho^i = \frac{m_0}{1 - \rho} < \infty,
\]

Hence get that

\[
\pi_i = \frac{m_i}{M} = (1 - \rho) \rho^i \quad \text{for all } i \geq 0.
\]

In other words, if \( \rho < 1 \) the invariant distribution is the geometric distribution with parameter \( \rho \).

Over a long time range the mean queue length is

\[
\sum_{i=1}^{\infty} i \pi_i = \sum_{i=1}^{\infty} (1 - \rho) i \rho^i = (1 - \rho) \rho \left( \sum_{i=1}^{\infty} \rho^i \right)' = \frac{\rho}{1 - \rho}.
\]
PS: In the case $\rho > 1$ the chain is not recurrent, in the case $\rho = 1$ the chain is null recurrent and no invariant distribution exists.

As a (more widely applicable) alternative one can also solve the equation $\pi Q = 0$ using the generating function

$$
\hat{\pi}(s) := \sum_{n=0}^{\infty} s^n \pi_n = \pi_0 + \pi_1 s + \pi_2 s^2 + \cdots.
$$

We write out the equation $\pi Q = 0$ as

$$
\begin{align*}
-\lambda \pi_0 + \mu \pi_1 &= 0 \\
\lambda \pi_0 - (\lambda + \mu) \pi_1 + \mu \pi_2 &= 0 \\
\lambda \pi_1 - (\lambda + \mu) \pi_2 + \mu \pi_3 &= 0 \\
&\quad \text{and so on. Multiplying the first equation by } s, \text{ the second by } s^2 \text{ and so forth and adding up, we get}
\end{align*}
$$

$$
\lambda s^2 \hat{\pi}(s) - (\lambda + \mu) s (\hat{\pi}(s) - \pi_0) + \mu (\hat{\pi}(s) - \pi_0) - \lambda \pi_0 s = 0,
$$

hence

$$
\hat{\pi}(s) (\lambda s^2 - (\lambda + \mu) s + \mu) = \pi_0 \mu (1 - s).
$$

This implies

$$
\hat{\pi}(s) = \frac{\pi_0 \mu (1 - s)}{\lambda s^2 - (\lambda + \mu) s + \mu} = \frac{\pi_0 \mu (1 - s)}{(s - 1)(\lambda s - \mu)} = \frac{\pi_0 \mu}{\mu - \lambda s}.
$$

To find $\pi_0$ recall that $\hat{\pi}(1) = 1$ hence

$$
\pi_0 = \frac{\mu - \lambda}{\mu} = 1 - \rho,
$$

and we infer that

$$
\hat{\pi}(s) = \frac{(1 - \rho) \mu}{\mu - \lambda s} = \frac{1 - \rho}{1 - \rho s}.
$$

To recover the $\pi_n$ we expand $\hat{\pi}(s)$ as a power series and equate the coefficients,

$$
\hat{\pi}(s) = (1 - \rho) \sum_{n=0}^{\infty} \rho^n s^n,
$$

hence $\pi_n = (1 - \rho) \rho^n$. The mean queue length in the long term can be found using

$$
\hat{\pi}'(s) = \sum_{n=1}^{\infty} n s^{n-1} \pi_n, \text{ and } \hat{\pi}'(s) = \frac{\rho (1 - \rho)}{(1 - \rho s)^2},
$$

which implies that the mean queue length is

$$
\sum_{n=1}^{\infty} n \pi_n = \hat{\pi}'(1) = \frac{\rho}{1 - \rho}.
$$
**Last Example:** The pure birth process.

Recall that the pure birth process $X$ is a continuous time Markov chain with $Q$-matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \ldots \\ 0 & -2\lambda & 2\lambda & 0 & \ldots \\ 0 & 0 & -3\lambda & 3\lambda & 0 & \ldots \\ 0 & \ldots & 0 & \ddots & \ddots & \ddots \\ \end{pmatrix}.$$ 

This process is strictly increasing, and hence transient, we show that, if $X(0) = 1$, then

$$\Pr\{X(t) = n\} = e^{-\lambda t}(1 - e^{-\lambda t})^{n-1} \text{ for } n = 1, 2, 3, \ldots, \tag{2.4.1}$$

i.e. $X(t)$ is geometrically distributed with parameter $e^{-\lambda t}$. Note that this shows that, as $t \to \infty$,

$$\Pr\{X(t) \leq n\} = \sum_{k=1}^{n} e^{-\lambda t}(1 - e^{-\lambda t})^{k-1} = 1 - (1 - e^{-\lambda t})^{n} \to 0,$$

hence the process $X(t)$ converges to infinity.

To prove (2.4.1) we use the resolvent method. We now write $\rho$ for the Laplace parameter, hence we have to invert

$$\rho I - Q = \begin{pmatrix} \rho + \lambda & -\lambda & 0 & 0 & \ldots \\ 0 & \rho + 2\lambda & -2\lambda & 0 & \ldots \\ 0 & 0 & \rho + 3\lambda & -3\lambda & 0 & \ldots \\ 0 & \ldots & 0 & \ddots & \ddots & \ddots \\ \end{pmatrix}.$$ 

The coefficients $r_{in}, n = 1, 2, \ldots$ of the inverse matrix hence satisfy

$$r_{11}(\rho + \lambda) = 1, \text{ and } r_{1,n-1}(-(n-1)\lambda) + r_{1,n}(\rho + n\lambda) = 0, \text{ for } n \geq 2.$$

Solving this gives

$$r_{1n} = \frac{(n-1)\lambda}{\rho + n\lambda} r_{1,n-1} = \lambda^{n-1}(n-1)! \prod_{k=1}^{n} \frac{1}{\rho + k\lambda}.$$

We use partial fractions, which means finding $a_k, k = 1, \ldots, n$ such that

$$r_{1n} = \frac{1}{\rho + k\lambda} = \sum_{k=1}^{n} \frac{a_k}{\rho + k\lambda}.$$

We thus get the equations

$$\lambda^{n-1}(n-1)! = \sum_{k=1}^{n} a_k \prod_{j \neq k} (\rho + j\lambda).$$

Plugging in $\rho = -\lambda k$ gives

$$\lambda^{n-1}(n-1)! = a_k \prod_{j \neq k} (\lambda(j-k)),$$

hence

$$a_k = \frac{(n-1)!}{\prod_{j \neq k} (j-k)} = \frac{(n-1)!}{(k-1)!} (-1)^{k-1}.$$
Now we have
\[
r_{1n} = \sum_{k=1}^{n} \frac{(n - 1)}{(k - 1)} (-1)^{k-1} \frac{1}{\rho + k\lambda}.
\]

Inverting the Laplace transform yields
\[
p_{1n}(t) = \sum_{k=1}^{n} \frac{(n - 1)}{(k - 1)} (-1)^{k-1} e^{-k\lambda t} = e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(n - 1)}{k} (-1)^{k} e^{-k\lambda t} = e^{-\lambda t} (1 - e^{-\lambda t})^{n-1},
\]
by the binomial formula. This verifies (2.4.1).