Continuous-time Markov chains

Continuous-Time Markov Chains

A stochastic process $(X_t)_{t\geq 0}$ in continuous time, taking values in a countable state space $\mathcal{S} \subset \mathbb{R}$ is said to be a Continuous-Time Markov Chain (CTMC) if, for all $k \geq 1$, $0 \leq t_1 < t_2 < \cdots < t_{k+1}$ and $i_1, i_2, \ldots, i_{k+1} \in \mathcal{S}$,

$$\begin{aligned} &\mathbb{P}(X_{t_{k+1}} = i_{k+1} | X_{t_1} = i_1, \cdots, X_{t_k} = i_k) \\ = \ &\mathbb{P}(X_{t_{k+1}} = i_{k+1} | X_{t_k} = i_k), \end{aligned}$$

whenever the left hand side is well defined.

As for DTMC, it is often convenient to assume that $\mathcal{S} = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, or that $\mathcal{S} = \mathbb{N}$.

If $\mathbb{P}(X_{t+s} = k | X_s = j)$ for $j, k \in \mathcal{S}$ do not depend on s then we say the CTMC is time homogeneous, and we can write $p_{j,k}^{(t)}$. We consider only time-homogeneous CTMCs in this course. By convention we assume that they are right continuous, i.e. $\mathbb{P}(\lim_{h\downarrow 0} X_{t+h} = X_t \text{ for all } t \geq 0) = 1$.

We can put the probabilities $p_{i,j}^{(t)}$ into a matrix $P^{(t)}$.

DTMC vs CTMC

For DTMC, if $X_n = i$ then we waited for a geometric$(1 - p_{i,i})$ amount of time before jumping to a *new state*. At the time we jump to a *new state*, the probability of jumping to j is $b_{i,j} = p_{i,j}/(1 - p_{i,i})$. If $p_{i,i} = 0$ then $b_{i,j} = p_{i,j}$.

For a CTMC, if $X_t = i$, we wait an exponential(λ_i) time and then jump to a new state. The probability of jumping to j is $b_{i,j}$.

This is equivalent to the following: let $(T_{i,j})_{j \in \mathcal{S}}$ be independent exponential$(q_{i,j})$ random variables. We wait time $T_i' = \min_{\ell \in \mathcal{S}} T_{i,\ell}$ at state i and then jump to the state k such that $T_{i,k} = T_i'$.

To see the equivalence, set $q_{i,j} = \lambda_i b_{i,j}$ then $T_i' \sim$exponential$(\sum_{\ell \in \mathcal{S}} q_{i,\ell})$ (where $\sum_{\ell \in \mathcal{S}} q_{i,\ell} = \lambda_i$) and the probability that we jump to j is

$$\mathbb{P}(T_{i,j} < \min_{\ell \neq j} T_{i,\ell}) = \frac{q_{i,j}}{\sum_{\ell \in \mathcal{S}} q_{i,\ell}} = b_{i,j}.$$

Transition diagrams

For a DTMC we drew a diagram containing the transition probabilities $p_{i,j}$.

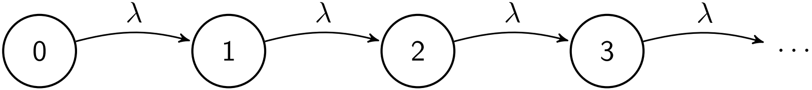
For a CTMC we draw a transition diagram containing the transition rates $q_{i,j}$.

The jump chain of a CTMC $(X_t)_{t\geq 0}$ is the DTMC $(X_n^J)_{n\in\mathbb{Z}_+}$ defined by $X_n^J = X_{T_n}$ where $T_0 = 0$, and $T_i = \inf\{t > T_{i-1} : X_t \neq X_{T_{i-1}}\}$ are the jump times of the CTMC. The transition probabilities of the jump chain are $b_{i,j}$. If $\lambda_i = 0$ we set $b_{i,i} = 1$ for the jump chain (otherwise $b_{i,i} = 0$).

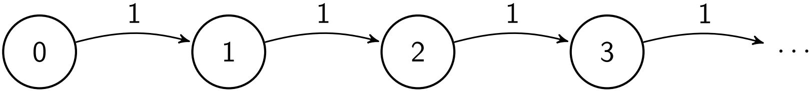
The transition diagram for the CTMC contains more information than that of its jump chain since the latter does not tell us how long the CTMC waits (on average) at each state.

Basic example: Poisson process

The transition diagram for the Poisson process $(N_t)_{t\geq 0}$ with rate λ is



The transition diagram for its jump process $(N_n^J)_{n\in\mathbb{Z}_+}$ is



Remarks:

Note that we have to wait for an *independent* exponential(λ_i) time on each successive visit to i

The waiting times must be exponential in order for the process to be memoryless/Markovian.

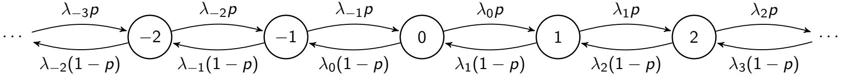
E.g. suppose $X_0 = j$ and T_1 is the first time the CTMC leaves j. Then the CTMC $(X_t)_{t\geq 0}$ must satisfy

$$\begin{aligned}
&\mathbb{P}(T_1 > t + s | T_1 > s) \\
&\quad = \mathbb{P}\big(X_v = j, \forall v \in [0, t+s] \big| X_u = j, \forall u \in [0, s]\big) \\
&\quad = \mathbb{P}\big(X_v = j, \forall v \in [s, t+s] \big| X_s = j\big) \text{ (Markov)} \\
&\quad = \mathbb{P}\big(X_v = j, \forall v \in [0, t] \big| X_0 = j\big) \text{ (homogeneous)} \\
&\quad = \mathbb{P}(T_1 > t)
\end{aligned}$$

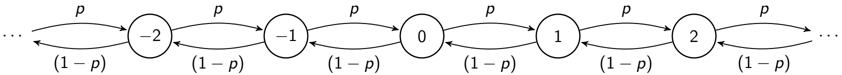
So T_1 must have an exponential distribution.

Example: continuous time random walk

Consider a CTMC with state space $\mathbb{Z}$ that waits for an exponential(λ_i) time in state $i \in \mathbb{Z}$ before jumping to $i+1$ with probability p and $i-1$ with probability $1-p$. Then the transition diagram is



The transition diagram of the jump chain is



If $\lambda_i = \lambda$ for each i we call the CTMC continuous time random walk.

Exercise:

Let $(X_t)_{t \geq 0}$ be a CTMC with state space $\mathcal{S} = \{1, 2, 3\}$ and transition rates $q_{i,j} = i$ for each i, j with $j \neq i$.

- Draw the transition diagram for $(X_t)_{t \geq 0}$
- Draw the transition diagram for the jump chain $(X_n^J)_{n \in \mathbb{Z}_+}$.

Classification of states and chains

As for DTMC we can ask about the following

(1) Whether a state i is absorbing (i.e. $\lambda_i = 0$)

(2) Whether $i \to j$ (i.e. $p_{i,j}^{(t)} > 0$ for some t)

(3) Communicating classes ($i \leftrightarrow j$ if $i \to j$ and $j \to i$)

(4) Irreducibility (i.e. $i \leftrightarrow j$ for every $i, j \in \mathcal{S}$)

(5) Hitting probabilities
(i.e. $h_{i,A} = \mathbb{P}(X_t \in A \text{ for some } t \geq 0 | X_0 = i)$)

(6) Recurrence (for irreducible chains: $h_{i,j} = 1$ for every $i, j \in \mathcal{S}$) and transience

(7) Expected hitting times (i.e. $m_{i,A}$, which is the expected time to reach a state in A starting from i)

(8) Positive recurrence (for irreducible chains: $m_{i,j} < \infty$ for every $i, j \in \mathcal{S}$)

(9) The long run behaviour of the chain (limiting proportion of time spent in state i etc.)

Some things are the same...

(1) A state i is an absorbing state for a CTMC if and only if $\lambda_i = 0$, if and only if $b_{i,i} = 1$, if and only if i is an absorbing state for the jump chain.

Similarly, items (2) to (6) above only depend on $(b_{i,j})_{i,j \in \mathcal{S}}$ and hence the CTMC has the given property if and only if its jump chain has that property.

On the other hand, items (7)-(9) depend on how long we wait (on average) at every state, so these properties will in general differ for the CTMC and its jump chain.

Exercise:

Consider a CTMC with state space $\mathcal{S} = \{1, 2, 3\}$ and transition rates $q_{2,1} = \lambda p$, $q_{2,3} = \lambda(1 - p)$, where $\lambda > 0$ and $p \in (0, 1)$, and all other transition rates are 0.

- Which states are absorbing?
- Find $m_{2,\{1,3\}}$.
- Find the hitting probability $h_{2,1}$.

Remarks

One can construct CTMC that are explosive in the sense that the process can jump infinitely many times in a finite amount of time. E.g. if $\mathcal{S} = \mathbb{Z}_+$ and $\lambda_i = \lambda^i$ for some $\lambda > 1$ and $b_{i,i+1} = 1$ for each i then the CTMC is explosive. We henceforth assume that our CTMC is not explosive.

A similar argument to the one we saw in the DTMC setting shows that an irreducible finite-state CTMC is positive recurrent.

Long run behaviour

As for DTMC, the limiting proportion of time spent by a CTMC in a given state can be random (recall our reducible examples).

For an irreducible transient CTMC, the limiting proportion of time spent in each state is 0, and there is no limiting *distribution*.

An irreducible positive recurrent CTMC is *ergodic*, i.e. the limiting distribution exists and does not depend on the initial distribution. The limiting distribution can be specified in terms of a quantity similar to the "expected return time" that appeared in DTMC. This is also equal to the stationary distribution of the CTMC (see next slide)

Stationary distribution

Recall that for a DTMC, a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is a stationary distribution for (a DTMC with transition matrix) P if $\pi P = \pi$.

This can be rewritten as $\pi(P - I) = 0$, where I is the $|\mathcal{S}| \times |\mathcal{S}|$ identity matrix.

A distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is called a *stationary distribution* for the family $(P^{(t)})_{t \geq 0}$ if $\pi P^{(t)} = \pi$ for each $t \geq 0$.
Let $q_{i,i} := -\sum_{j \neq i} q_{i,j} = -\lambda_i$ and let Q denote the matrix (called the (infinitesimal) generator, or rate matrix) whose i, jth entry is $q_{i,j}$.

For non-explosive CTMCs a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ is a stationary distribution for (a Markov chain with rate matrix) Q if and only if $\pi Q = 0$.

This is equivalent to the set of equations $\pi_i \lambda_i = \sum_{j \neq i} \pi_j q_{j,i}$ for $i \in \mathcal{S}$ which are referred to as the full balance equations.

The main result

Theorem: An irreducible and positive recurrent CTMC has a unique stationary distribution π. For such a CTMC, the limiting proportion of time spent in state i is π_i and the limiting distribution is π (irrespective of the initial distribution).

Note that periodicity is not an issue for a CTMC.

For $i \in \mathcal{S}$, let $T_1^{(i)} = \inf\{t > 0 : X_t \neq i\}$ and $T^{i,i} = \inf\{t > T_1^{(i)} : X_t = i\}$. Then

$$\pi_i = \frac{\mathbb{E}[T_1^{(i)} | X_0 = i]}{\mathbb{E}[T^{i,i} | X_0 = i]}.$$

The stationary distribution continued

The quantity $\frac{\mathbb{E}[T_1^{(i)}|X_0=i]}{\mathbb{E}[T^{i,i}|X_0=i]}$ is a bit like the proportion of time spent at i up to the first time that we *return* to i (start from i initially).

The numerator of this quantity is $\frac{1}{\lambda_i} = -\frac{1}{q_{i,i}}$. By a first step analysis, the denominator is

$$\frac{1}{\lambda_i} + \sum_{j \in \mathcal{S}} b_{i,j} m_{j,i}.$$

This is because on average we take time $1/\lambda_i$ to escape from i, at which point we go to j with probability $b_{i,j}$ and then we have to get from j to i (which takes time $m_{j,i}$ on average).

Reversibility

Suppose that we find a distribution $\pi = (\pi_i)_{i \in \mathcal{S}}$ such that

$$\pi_i q_{i,j} = \pi_j q_{j,i}, \qquad \text{for all } i, j \in \mathcal{S}.$$

Then we say that Q is reversible.

The above equations are called the detailed balance equations.

Note that (exercise!) if π satisfies the detailed balance equations then it satisfies the full balance equations too. Thus any distribution satisfying the detailed balance equations is a stationary distribution.

Explosive CTMC revisited

For explosive CTMCs, it is possible to have a solution to

$$\boldsymbol{\pi} Q = 0,$$

with $\sum_j \pi_j = 1$ that is not the stationary distribution.

- Take the CTMC with $q_{i,i+1} = \lambda_i p$ for $i \geq 0$, and $q_{i,i-1} = (1-p)\lambda_i$ for $i \geq 1$ as the only non-zero transition rates.
- If $p > 1 - p$, then the chain is transient, but choosing $\lambda_i = \lambda^i$ there is a solution to
$$\boldsymbol{\pi} Q = 0$$
of the form $\pi_i = \pi_0 \left(\frac{p}{(1-p)\lambda}\right)^i$. Thus we can get a solution that sums to 1 if $\lambda > p/(1-p)$.

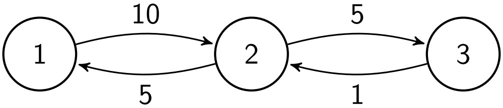
Expected hitting times

Theorem: For a CTMC with state space $\mathcal{S}$, and $A \subset \mathcal{S}$, the vector of mean hitting times $(m_{i,A})_{i \in \mathcal{S}}$ is the minimal non-negative solution to

$$m_{i,A} = \begin{cases} 0, & \text{if } i \in A, \\ \frac{1}{\lambda_i} + \sum_{j \in \mathcal{S}} b_{i,j} m_{j,A}, & \text{if } i \notin A. \end{cases}$$

Exercise:

Let $(X_t)_{t\geq 0}$ be a CTMC with the following transition diagram



- Find $m_{1,3}$.
- Find the stationary distribution for this chain.
- Find the stationary distribution for the jump chain.

The Chapman-Kolmogorov equations

Observe that

$$\begin{aligned} p_{i,j}^{(s+t)} &= \sum_k \mathbb{P}(X_{s+t} = j | X_s = k, X_0 = i)\mathbb{P}(X_s = k | X_0 = i) \\ &= \sum_k p_{i,k}^{(s)} p_{k,j}^{(t)}. \end{aligned}$$

These are the Chapman-Kolmogorov equations for a CTMC. In matrix form, the Chapman-Kolmogorov equations can be expressed as

$$P^{(t+s)} = P^{(s)} P^{(t)}.$$

Finding the transition probabilities

Thus far we have not actually computed $P^{(t)}$.

By analogy with the discrete-time case, we might hope that we can write $P^{(t)} = P^t$ for some matrix P.

If $t = m$ (a positive integer), the C-K equations tell us that $P^{(m)} = (P^{(1)})^m$ and our hope is fulfilled, but if $t < 1$?

We want a single object (like $P = P^{(1)}$ in the discrete case) that encodes the information of the chain.

It turns out that the generator Q is our single object. In fact we have the so-called Forward and Backward equations:

$$\frac{d}{dt} P^{(t)} = QP^{(t)} \quad (\ddagger) \qquad \text{backward equations}$$

and,

$$\frac{d}{dt} P^{(t)} = P^{(t)} Q. \quad (\dagger) \qquad \text{forward equations}$$

Solving the forward and backward equations

For (non-explosive) CTMCs, the matrix Q determines the transition probability completely by solving the backward or forward equations to get

$$
\begin{aligned}
P^{(t)} &= \exp(tQ) \\
&:= \sum_{k=0}^{\infty} \frac{1}{k!} t^k Q^k,
\end{aligned}
$$

subject to $P^{(0)} = I$.

Example: The Poisson process

The Poisson process is a CTMC with generator

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

Poisson process transition probabilities

Can we derive the transition probabilities from Q? We could

- Compute $\exp(tQ)$.
- Solve the Kolmogorov backward or forward differential equations.

For the first case, one can show that $(Q^n)_{i,j} = 0$ if $j \notin \{i, i+1, \ldots, i+n\}$, and otherwise

$$(Q^n)_{i,j} = \lambda^n \binom{n}{j-i} (-1)^{n-(j-i)}.$$

Then for $j \geq i$,

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} (Q^n)_{i,j} = \frac{(t\lambda)^{j-i}}{(j-i)!} e^{-t\lambda}.$$

Or we can directly solve e.g. the forward equations to find $p_{0,k}^{(t)}$.

Poisson process transition probabilities

Now,

$$\begin{aligned} \frac{d}{dt} p_{0,0}^{(t)} &= -\lambda p_{0,0}^{(t)} \\ \implies p_{0,0}^{(t)} &= c e^{-\lambda t}. \end{aligned}$$

So with the condition that $p_{0,0}^{(0)} = 1$ we get $p_{0,0}^{(t)} = e^{-\lambda t}$.Similarly

$$\frac{d}{dt} p_{0,k}^{(t)} = \sum_{j=0}^{k} p_{0,j}^{(t)} q_{j,k} = \lambda (p_{0,k-1}^{(t)} - p_{0,k}^{(t)})$$

By induction, we can show that $p_{0,k}^{(t)} = e^{-\lambda t} (\lambda t)^k / k!$.

Interpretation of the generator

For small h,

$$\begin{array}{rcl} \mathbb{P}(X_{t+h} = k | X_t = j) & = & p_{j,k}^{(h)} \\ & \approx & (I + hQ)_{j,k} \\ & = & \left\{ \begin{array}{ll} hq_{j,k}, & \text{if } j \neq k, \\ 1 + hq_{j,j}, & \text{if } j = k. \end{array} \right. \end{array}$$

So indeed for $k \neq j$ we can think of $q_{j,k}$ as the rate of transition from j to k.

Example - birth and death processes

Let $(X_t)_{t \geq 0}$ be a CTMC on $\mathcal{S} = \mathbb{Z}_+$, where:

- X_t represents the number of 'people' in a system at time t.
- Whenever there are n 'people' in the system
  - new arrivals enter (by birth or immigration) the system at rate ν_n
  - 'people' leave (or die from) the system at rate μ_n
  - arrivals and departures occur independently of one another
- $(X_t)_{t \geq 0}$ is a birth-and-death process with arrival (or birth) rates $(\nu_n)_{n \in \mathcal{S}}$ and departure (or death) rates $(\mu_n)_{n \in \mathcal{S}}$.

Generator of a birth and death process

The generator of such a birth and death process has the form

$$Q = \begin{pmatrix} -\nu_0 & \nu_0 & 0 & 0 & 0 & \ddots \\ \mu_1 & -(\mu_1+\nu_1) & \nu_1 & 0 & 0 & \ddots \\ 0 & \mu_2 & -(\mu_2+\nu_2) & \nu_2 & 0 & \ddots \\ 0 & 0 & \mu_3 & -(\mu_3+\nu_3) & \nu_3 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

The CTMC evolves by remaining in state k for an exponentially-distributed time with rate $\nu_k + \mu_k$, then it moves to state $k+1$ with probability $b_{k,k+1} = \nu_k/(\nu_k+\mu_k)$ and state $k-1$ with probability $b_{k,k-1} = \mu_k/(\nu_k+\mu_k)$, and so on.

Example

Consider a population in which there are no births ($\nu_i = 0$ for $i \geq 0$) and, for $i \geq 1$, the death rates are μ_i.
Suppose X_t is the population size at time t and so $(X_t)_{t \geq 0}$ is a CTMC.

1. Find expressions for $p_{i,i}^{(t)}$ and $p_{i,i-1}^{(t)}$ for $i \geq 1$.
2. Given that the population size is two at a particular time, calculate the probability that no more than one death will occur within the next one unit of time.

Birth and death stationary distribution

Assume $\nu_i > 0$ and $\mu_i > 0$ for all i. Assuming non-explosivity, we derive the stationary distribution (if it exists) by solving $\boldsymbol{\pi} Q = 0$. In fact we can solve the detailed balance equations:

$$\nu_k \pi_k = \mu_{k+1} \pi_{k+1}, \qquad k \in \mathbb{Z}_+$$

This has solution

$$\pi_k = \pi_0 \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell}.$$

So a stationary distribution exists if and only if

$$\sum_{k=0}^{\infty} \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell} < \infty$$

in which case

$$\pi_0 = \left(\sum_{k=0}^{\infty} \prod_{\ell=1}^{k} \frac{\nu_{\ell-1}}{\mu_\ell} \right)^{-1}.$$

Exercises

- Compute the stationary distribution of an 'M/M/1 queue', which is a birth and death process with $\nu_i = \nu$, $i \geq 0$, and $\mu_i = \mu$, $i \geq 1$.
- Compute the stationary distribution of a birth and death process with constant birth rate $\nu_i = \nu$, $i \geq 0$, and unit per capita death rate $\mu_i = i$, $i \geq 1$.

Finding the transition probabilities - the details

If t and h are nonnegative real numbers, we can write

$$\begin{aligned} \frac{P^{(t+h)} - P^{(t)}}{h} &= P^{(t)} \left[\frac{P^{(h)} - I}{h} \right] \\ &= \left[\frac{P^{(h)} - I}{h} \right] P^{(t)} \end{aligned}$$

This suggests that we should investigate the existence of the derivative

$$Q^* \equiv \lim_{h \to 0^+} \frac{P^{(h)} - I}{h}.$$

Under our assumptions about the chain (non-explosive, etc.), Q^* exists and equals Q.

Forward and backward equations

Since

$$\begin{aligned}\frac{P^{(t+h)} - P^{(t)}}{h} &= P^{(t)} \left[\frac{P^{(h)} - I}{h}\right] \\ &= \left[\frac{P^{(h)} - I}{h}\right] P^{(t)}\end{aligned}$$

if we can take the limits through the matrix multiplication then we get

$$\frac{d}{dt} P^{(t)} = Q P^{(t)} \quad (\ddagger) \qquad \text{backward equations}$$

and, similarly,

$$\frac{d}{dt} P^{(t)} = P^{(t)} Q. \quad (\dagger) \qquad \text{forward equations}$$

Why does $Q^* = Q$?

Let T_1 denote the time of the first jump of our process and T_2 denote the time of the second jump.

Then

$$\mathbb{P}(T_1 > h | X_0 = i) = e^{-\lambda_i h} = 1 - \lambda_i h + o(h) = 1 + q_{i,i} h + o(h).$$

$$\begin{aligned}
\mathbb{P}(T_2 < h | X_0 = i) &= \sum_{k \in \mathcal{S}} \mathbb{P}(T_1 < h, T_2 - T_1 < h - T_1, X_{T_1} = k | X_0 = i) \\
&\leq \sum_{k \in \mathcal{S}} \mathbb{P}(T_1 < h, T_2 - T_1 < h, X_{T_1} = k | X_0 = i) \\
&= \sum_{k \in \mathcal{S}} (1 - e^{-\lambda_k h})(1 - e^{-\lambda_i h}) b_{i,k} \\
&= \sum_{k \in \mathcal{S}} (\lambda_k h + o(h))(\lambda_i h + o(h)) b_{i,k} \\
&= h^2 \sum_{k \in \mathcal{S}} (\lambda_k + o(1))(\lambda_i + o(1)) b_{i,k} = o(h),
\end{aligned}$$

provided that λ_k do not grow too fast with k.

Why does $Q^* = Q$?

So, $\mathbb{P}(T_1 > h | X_0 = i) = 1 - \lambda_i h + o(h)$. Therefore $\mathbb{P}(T_1 < h | X_0 = i) = \lambda_i h + o(h)$.
Also $\mathbb{P}(X_{T_1} = j | X_0 = i) = b_{i,j} = q_{i,j}/\lambda_i$.
Also $\mathbb{P}(T_2 < h | X_0 = i) = o(h)$.
So

$$p_{i,i}^{(h)} = \mathbb{P}(T_1 > h | X_0 = i) + \mathbb{P}(T_2 < h, X_h = i | X_0 = i) = 1 - \lambda_i h + o(h).$$

Similarly, for $j \neq i$,

$$\begin{aligned} p_{i,j}^{(h)} &= \mathbb{P}(X_{T_1} = j, T_1 < h, T_2 > h | X_0 = i) + \mathbb{P}(T_2 < h, X_h = j | X_0 = i) \\ &= \frac{q_{i,j}}{\lambda_i} \lambda_i h + o(h). \end{aligned}$$

Now you can see that $(P^{(h)} - I)_{i,j} = q_{i,j} h + o(h)$. Divide by h and take the limit...

Justifying the forward and backward equations

We write "hope that" since we need to justify pushing the limits through the (possibly infinite) sums

$$\lim_{h\to 0}\sum_{k\in\mathcal{S}} p_{i,k}^{(t)}\left[\frac{P^{(h)}-I}{h}\right]_{k,j} = \sum_{k\in\mathcal{S}} p_{i,k}^{(t)}\lim_{h\to 0}\left[\frac{P^{(h)}-I}{h}\right]_{k,j}$$

$$\lim_{h\to 0}\sum_{k\in\mathcal{S}}\left[\frac{P^{(h)}-I}{h}\right]_{i,k} p_{k,j}^{(t)} = \sum_{k\in\mathcal{S}}\lim_{h\to 0}\left[\frac{P^{(h)}-I}{h}\right]_{i,k} p_{k,j}^{(t)}$$

for each $i, j \in S$.

If $\mathcal{S}$ is finite, then there is no problem and both (‡) and (†) hold.

Justifying the backward equations

In fact, we know from Fatou's Lemma that, for $j, k \in S$,

$$\begin{aligned}
\liminf_{h \to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} &= \liminf_{h \to 0^+} \sum_{i \in S} \frac{(p_{ji}^{(h)} - \delta_{ji}) p_{ik}^{(t)}}{h} \\
&\geq \sum_{i \in S} \lim_{h \to 0^+} \frac{(p_{ji}^{(h)} - \delta_{ji}) p_{ik}^{(t)}}{h} \\
&= \sum_{i \in S} a_{ji} p_{ik}^{(t)}.
\end{aligned}$$

Similarly,

$$\liminf_{h \to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} \geq \sum_{i \in S} p_{ji}^{(t)} a_{ik}.$$

Justifying the backward equations

We can show that the inequality in the first expression is, in fact, an equality, as follows. For $N > j$,

$$
\begin{aligned}
\sum_{i\in S} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} &= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \sum_{i=N+1}^{\infty} \frac{p_{ji}^{(h)} p_{ik}^{(t)}}{h} \\
&\leq \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \sum_{i=N+1}^{\infty} \frac{p_{ji}^{(h)}}{h} \\
&= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} + \frac{1 - \sum_{i=1}^{N} p_{ji}^{(h)}}{h} \\
&= \sum_{i=1}^{N} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] \left[p_{ik}^{(t)} - 1\right]}{h}.
\end{aligned}
$$

Justifying the backward equations

Therefore

$$\begin{aligned}
\limsup_{h\to 0^+} \sum_{i\in S} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] p_{ik}^{(t)}}{h} &\leq \sum_{i=1}^{N} \lim_{h\to 0^+} \frac{\left[p_{ji}^{(h)} - \delta_{ji}\right] \left[p_{ik}^{(t)} - 1\right]}{h} \\
&= \sum_{i=1}^{N} a_{ji} p_{ik}^{(t)} - \sum_{i=1}^{N} a_{ji}.
\end{aligned}$$

Now we let $N \to \infty$ and use the fact that $\sum_{i=1}^{\infty} a_{ji} = 0$ to derive

$$\limsup_{h\to 0^+} \frac{p_{jk}^{(t+h)} - p_{jk}^{(t)}}{h} \leq \sum_{i\in S} a_{ji} p_{ik}^{(t)},$$

which proves that $p_{jk}^{(t)}$ is differentiable (since $\liminf \geq \limsup$) and

$$\frac{dp_{jk}^{(t)}}{dt} = \sum_{i\in S} a_{ji} p_{ik}^{(t)}.$$

Justifying the forward and backward equations

So we have justified the interchange leading to (‡). However, the interchange leading to (†) does not hold for explosive CTMCs.