Continuous Time Markov Chains

Define

$$h_{jk}(t',t'') \equiv \Pr\left[S(t'') = k \middle| S(t') = j\right],\ t'' \geq t'.$$
$$\underline{\underline{H}}(t',t'') \equiv \left[h_{jk}(t',t'')\right]$$

For a homogeneous Markov chain, the transition probabilities are functions only of the difference $t'' - t'$:

$$\underline{\underline{H}}(t',t'') = \underline{\underline{H}}(0,t''-t')\ \ \forall t',t''\ s.t.\ t'' \geq t' \geq 0.$$

All of the Markov chains we consider will be homogeneous, unless stated otherwise.

Assume that $t'' - t' = \Delta t$ is small.

The probability of n transitions (events or steps) in Δt is proportional to $(\Delta t)^n$ for small Δt. To see this, recall that in a continuous time Markov chain, the time between state transitions is exponentially distributed. This implies that the transition times are generated by a Poisson process.

A Poisson process can be defined in several ways. We will use the following set of axioms:

1. $\Pr[1$ event in an interval of length $\Delta t \to 0] = \lambda \Delta t$.
2. $\Pr[0$ events in an interval of length $\Delta t \to 0] = 1 - \lambda \Delta t - o\left((\Delta t)^2\right)$.
3. Events are independent.

The first axiom says that the probability of one event occurring in a very short interval is proportional to the length of the interval. The second axiom states that the probability of no Poisson events occurring during a very short interval is one minus the probability of one event minus a term which is $o\left((\Delta t)^2\right)$. The "little o" notation means that $\lim_{\Delta t \to 0} o\left((\Delta t)^2\right) \big/ \Delta t = 0$.

Define $P_i(t) \equiv \Pr[i$ events in an interval of length $t]$. Note that here t may be arbitrarily large. We can determine these probabilities for all values of i by starting at $i = 0$ and working up.

$$\begin{aligned} P_0(t+\Delta t) &= P_0(t) \cdot P_0(\Delta t) \\ &= P_0(t)\left[1 - \lambda \Delta t - o\left((\Delta t)^2\right)\right] \end{aligned}$$

The last substitution relies on Δt being very small. Manipulating this equation and dividing both sides by Δt,

$$\frac{P_0(t+\Delta t)-P_0(t)}{\Delta t}=-\lambda P_0(t)-\frac{o\left((\Delta t)^2\right)}{\Delta t}P_0(t)$$

Taking the limit of both sides of the equation as $\Delta t \to 0$, we get

$$\frac{d}{dt}\{P_0(t)\}=-\lambda P_0(t)$$

The solution to this linear, first-order, time-invariant differential equation is

$$P_0(t)=ke^{-\lambda t}$$

for some constant k. To determine k, note that the probability of 0 events in 0 time is 1:

$$P_0(0)=1=ke^{-\lambda\cdot 0}=k$$

Now consider $i=1$.

$$\begin{aligned}P_1(t+\Delta t)&=P_1(t)\cdot P_0(\Delta t)+P_0(t)\cdot P_1(\Delta t)\\&=P_1(t)[1-\lambda\Delta t]+e^{-\lambda t}\cdot\lambda\Delta t\end{aligned}$$

Note that we have dropped the $o\left((\Delta t)^2\right)$ from the expression for $P_0(\Delta t)$. We are going to be playing the same game as before (dividing by Δt and taking the limit as $\Delta t \to 0$), and the $o\left((\Delta t)^2\right)$ will disappear anyway.

$$\begin{aligned}\frac{P_1(t+\Delta t)-P_1(t)}{\Delta t}&=\lambda e^{-\lambda t}-\lambda P_1(t)\\\lim_{\Delta t\to 0}\frac{P_1(t+\Delta t)-P_1(t)}{\Delta t}&=\lambda e^{-\lambda t}-\lambda P_1(t)\\\frac{d}{dt}\{P_1(t)\}&=\lambda e^{-\lambda t}-\lambda P_1(t)\end{aligned}$$

which has the solution

$$P_1(t)=\lambda te^{-\lambda t}$$

In general, for $i\geq 1$,

$$P_i(t+\Delta t) = P_i(t)\cdot P_0(\Delta t) + P_{i-1}(t)\cdot P_1(\Delta t)$$
$$= P_i(t)[1-\lambda\Delta t] + P_{i-1}(t)\cdot \lambda\Delta t$$
$$\frac{d}{dt}\{P_i(t)\} = \lambda P_{i-1}(t) - \lambda P_i(t)$$

$$P_i(t) = \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

You can verify this solution by substituting in the differential equation.

This shows that the probability of i events in an interval of length t has a *Poisson distribution*, and λ is the *rate* or the *parameter* of the distribution. Accordingly, the probability of n events in Δt is proportional to $(\Delta t)^n$ for small Δt (just use a power series expansion for $e^{-\lambda\Delta t}$).

To show the connection between the Poisson process and the exponential distribution, we need to prove that the time between Poisson events is exponentially distributed. Let X be the random variable for the time between two consecutive events of a Poisson process.

$$\Pr[X > t] = \Pr[0 \text{ events in } t] = P_0(t) = e^{-\lambda t}$$
$$= 1 - \Pr[X \le t]$$
$$\Pr[X \le t] = 1 - e^{-\lambda t} = F_X(t)$$

where $F_X(t)$ is the cumulative distribution function for an exponentially distributed random variable X.

Returning to the discussion of homogeneous Markov chains, let $\underline{p}(t)$ be the state probability vector at time t. Because the probability of n events in Δt is proportional to $(\Delta t)^n$ for small Δt, the probability of 1 event dominates as Δt gets close to 0 (not surprising - it is a Poisson process, after all). Hence, as $\Delta t \to 0$, $\lim_{\Delta t \to 0} \underline{\underline{H}}(t, t+\Delta t) = \lim_{\Delta t \to 0} \underline{\underline{H}}(0, \Delta t) \equiv \underline{\underline{P}}(\Delta t)$ can be thought of as the single-step transition matrix for the continuous time Markov chain. Then

$$\lim_{\Delta t \to 0} \underline{p}(t) \cdot \underline{\underline{H}}(t, t+\Delta t) = \lim_{\Delta t \to 0} \underline{p}(t+\Delta t)$$

$$\lim_{\Delta t \to 0} \frac{\underline{p}(t) \cdot \underline{\underline{H}}(t, t+\Delta t) - \underline{p}(t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\underline{p}(t+\Delta t) - \underline{p}(t)}{\Delta t}$$

$$\underline{p}(t) \cdot \lim_{\Delta t \to 0} \frac{\underline{\underline{H}}(t, t+\Delta t) - \underline{\underline{I}}}{\Delta t} = \frac{d}{dt}\left\{\underline{p}(t)\right\}$$

$$\underline{p}(t) \cdot \underline{\underline{Q}}(t) = \frac{d}{dt}\left\{\underline{p}(t)\right\}$$

where $\underline{\underline{Q}}(t) = \lim_{\Delta t \to 0} \frac{\underline{\underline{H}}(t, t+\Delta t) - \underline{\underline{I}}}{\Delta t}$ is the *transition rate matrix* or *rate generator matrix* or simply the *generator matrix* of the continuous time Markov chain. The off-diagonal elements of $\underline{\underline{Q}}(t)$ are

$$q_{jk}(t) = \lim_{\Delta t \to 0} \frac{h_{jk}(t, t+\Delta t)}{\Delta t}, \quad j \neq k$$

are *not* probabilities; they are instantaneous rates of change in probability. Because the chain is homogeneous and must be memoryless,

$$\lim_{\Delta t \to 0} h_{jk}(t, t+\Delta t) = \lambda_{jk} \Delta t$$

and hence

$$q_{jk}(t) = q_{jk} = \lambda_{jk}$$

λ_{jk} is the rate of the Poisson process governing transitions from state j to state k. What are the $q_{jj}(t)$s, the diagonal elements of the generator matrix?

$$q_{jj}(t) = \lim_{\Delta t \to 0} \frac{h_{jj}(t, t+\Delta t) - 1}{\Delta t}$$

$h_{jj}(t, t+\Delta t)$ is the probability that the Markov chain is in state j at time $t+\Delta t$, given that it was in state j at time t. Since the chain must be in *some* state at time $t+\Delta t$, given that it was in state j at time t,

$$h_{jj}(t, t+\Delta t) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} h_{jk}(t, t+\Delta t)$$

and

$$\lim_{\Delta t \to 0} h_{jj}(t, t+\Delta t) = 1 - \lim_{\Delta t \to 0} \sum_{\substack{j=1 \\ j \neq k}}^{n} h_{jk}(t, t+\Delta t)$$

$$= 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_{jk} \Delta t = 1 - \lambda_j \Delta t$$

where $\lambda_j = \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_{jk}$ is the sum of the rates of the Poisson processes governing the transitions *out* of state j. Substituting this into the expression for $q_{jj}(t)$ and again using the homogeneity of the Markov chain, we get

$$q_{jj}(t) = q_{jj} = \lim_{\Delta t \to 0} \frac{1 - \lambda_j \Delta t - 1}{\Delta t} = -\lambda_j$$

Take a moment to consider λ_j. Let $\{X_1, \ldots, X_n\}$ be a set of independent, exponentially distributed random variables with rates $\lambda_1, \ldots, \lambda_n$, and let $X \equiv \min\{X_1, \ldots, X_n\}$.

$$\begin{aligned}\Pr[X > t] &= \Pr[X_1 > t \,\&\, X_2 > t \,\&\ldots\&\, X_n > t] \\ &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \cdot \ldots \cdot \Pr[X_n > t] \\ &= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} \cdot \ldots \cdot e^{-\lambda_n t} = e^{-\lambda t}\end{aligned}$$

where $\lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. Hence X is also exponentially distributed, with parameter (rate) λ. This says that λ_j is the rate of an exponentially distributed random variable that is the minimum of the random variables representing the times until transitions from state j to all other states. That is, λ_j is the rate of an exponentially distributed random variable that represents the time spent in state j. One way to look at it is that λ_j is the rate at which probability mass "leaves" state j.

We're almost there. So far, we have

$$\underline{p}(t) \cdot \underline{\underline{Q}} = \frac{d}{dt}\{\underline{p}(t)\}$$

We are interested in the steady-state probability vector $\underline{\pi} = \lim_{t \to \infty} \underline{p}(t)$.

$$\underline{\pi} \cdot \underline{\underline{Q}} = \lim_{t \to \infty} \frac{d}{dt}\{\underline{p}(t)\} = \frac{d}{dt}\left\{\lim_{t \to \infty} \underline{p}(t)\right\} = \frac{d}{dt}\{\underline{\pi}\} = \underline{0}$$

since the derivative of the steady-state probability vector is by definition the 0 vector.

The matrix equation $\underline{\pi} \cdot \underline{\underline{Q}} = \underline{0}$ for continuous time Markov chains is the analog of $\underline{\pi} \cdot \underline{\underline{P}} = \underline{\pi}$ for discrete time Markov chains. The two matrices are quite different. The elements of $\underline{\underline{P}}$ are probabilities; the elements of $\underline{\underline{Q}}$ are rates of change in probability. However, both matrices are singular, since each row of $\underline{\underline{P}}$ sums to 1 and each row of $\underline{\underline{Q}}$ sums to 0.