Statistical Methods in Bioinformatics, Univ. of Makerere
Markov chains in continuous time
Timo Koski 2018-08-02

1 Introduction

This lecture deals with stochastic processes known as Markov chains in continuous time. These processes will in the next lecture describe the substitution process on a sequence position. The lecture corresponds to sections 4.1, 10.7, and 13.3.1 in Ewens and Grant.

2 Definition and some first properties

Let $X = \{X(t) \mid t \geq 0\}$ be a family of random variables taking values in a discrete, countable, alphabet or state space $\mathcal{X}$. The variable t is called time. We denote the generic elements of $\mathcal{X}$ by $j, i, \ldots,$. The special case we have in mind is the finite state space $\mathcal{X} = \{A, T, C, G\}$.

2.1 The Markov property

The process $X = \{X(t) \mid t \geq 0\}$ is called a continuous-time Markov chain if it satisfies the following definition.

Definition 2.1 $X = \{X(t) \mid t \geq 0\}$ *satisfies the* **Markov property**, *if*

$$P\left(X(t_n) = j | X(t_1), X(t_2), \ldots, X(t_{n-1})\right) = P\left(X(t_n) = j | X(t_{n-1})\right) \quad (2.1)$$

for $j \in \mathcal{X}$, $i \in \mathcal{X}$, *and any sequence* $t_1 < t_2 < \ldots < t_{n-1} < t_n$ *of times.*

■

The evolution of continuous-time Markov chains can be described in very much the same terms as those used for Markov Chains.

The general situation is as follows. For Markov Chains we wrote the n-step transition probabilities in matrix form and expressed them in terms of the one-step matrix $\mathbf{P}$. In continuous time there is no analogue for $\mathbf{P}$, since there is no implicit unit length of time. Some differential calculus enables us to see that there is a matrix $\mathbf{Q}$, called the **generator** of the continuous-time chain, which takes over the role of $\mathbf{P}$.

2.2 The transition probability

Definition 2.2 *The* **time-homogeneous transition probability** *is denoted by $P_{ij}(t)$ and is defined as*

$$P_{ij}(t) = P\left(X(t) = j | X(0) = i\right) \tag{2.2}$$

or

$$P_{ij}(t-s) = P\left(X(t) = j | X(s) = i\right) \tag{2.3}$$

for $j \in \mathcal{X}$.

■

This is most readily presented in a matrix form.

$$\mathbf{P}(t) = \{P_{ij}(t)\}_{i \in \mathcal{X}, j \in \mathcal{X}}.$$

2.3 Chapman - Kolmogorov equations

Proposition 2.1 *The family $\{\mathbf{P}(t) | t \geq 0\}$ satisfies*

(a) $\mathbf{P}(0) = \mathbf{I}$ *(= the identity matrix).*

(b) $\mathbf{P}(t)$ *is a stochastic matrix.*

(c) the Chapman - Kolmogorov equations

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s). \tag{2.4}$$

■

Proof:

(a) This is obvious by the definitions.

(b) Clearly

$$P_{ij}(t) \geq 0. \tag{2.5}$$

Also, since the events $\{X(t) = j\}$ are disjoint,

$$\sum_{j \in \mathcal{X}} P_{ij}(t) = P\left(\cup_{j \in \mathcal{X}} \{X(t) = j\} | X(0) = i\right) = 1 \tag{2.6}$$

Hence $\mathbf{P}(t)$ is a stochastic matrix.

(c)

$$P_{ik}(t+s) = P\left(X(t+s)=j|X(0)=i\right) =$$
$$= \sum_{k\in\mathcal{X}} P\left(X(t+s)=j|X(s)=k\right)P\left(X(s)=k|X(0)=i\right)$$
$$= \sum_{k\in\mathcal{X}} P_{ik}(t)P_{kj}(s), \tag{2.7}$$

where we used the Markov property (2.1). The equality is in matrix form written as

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s). \tag{2.8}$$

By this the proof is completed. ∎

Most questions about $X = \{X(t) \mid t \geq 0\}$ can be answered in terms of the matrices $\mathbf{P}(t)$.

Assumption 2.1 *We shall now assume that the transition probabilities $P_{ij}(t)$ are continuous functions of t. We shall also assume that*

$$\mathbf{P}(t) \to \mathbf{I}, \quad \text{as } t \downarrow 0. \tag{2.9}$$

∎

This is to say that

$$P_{ij}(t) \to 0, i \neq j, \quad P_{ii}(t) \to 1, i = j, \quad \text{as } t \downarrow 0.$$

3 The generator

We make another assumption.

Assumption 3.1 *We assume that*

$$P_{ij}(h) = q_{ij}h + o(h), \quad i \neq j, \tag{3.1}$$

and

$$P_{ii}(h) = 1 + q_{ii}h + o(h), \tag{3.2}$$

where $o(h)$ ('small ordo') is a function such that $o(h)/h \to 0$, as $h \to 0$.

■

The numbers q_{ij} are known as the *instantaneous transition rates* or *intensities* of the continuous-time Markov chain. We are here assuming that the probability of two or more transitions in an interval $t, t+h$ is small. This can in fact be proved in a more rigorous treatment. Note that we are also implicitly thinking that the transition rates are not infinite.

From (3.1) and (3.2) we get

$$1 = \sum_{j \in \mathcal{X}} P_{ij}(h) = h \sum_{j \in \mathcal{X}, j \neq i} q_{ij} + 1 + q_{ii}h + o(h), \tag{3.3}$$

and this implies

$$h \sum_{j \in \mathcal{X}, j \neq i} q_{ij} = -q_{ii}h + o(h)$$

or, by dividing by h and letting h go to zero,

$$\sum_{j \in \mathcal{X}, j \neq i} q_{ij} = -q_{ii} \tag{3.4}$$

or

$$\sum_{j \in \mathcal{X}} q_{ij} = 0 \tag{3.5}$$

We introduce the symbol q_i by

$$q_i \stackrel{\text{def}}{=} \sum_{j \in \mathcal{X}, j \neq i} q_{ij}. \tag{3.6}$$

The assumption (3.2) gives thus

$$\lim_{h \downarrow 0} \frac{P_{ii}(h) - 1}{h} = q_{ii} = -q_i, \tag{3.7}$$

and the assumption (3.2)

$$\lim_{h \downarrow 0} \frac{P_{ij}(h)}{h} = q_{ij}, i \neq j. \tag{3.8}$$

We introduce the square matrix

$$\mathbf{Q} = (q_{ij})_{i,j \in \mathcal{X}}. \tag{3.9}$$

In matrix form (3.7) and (3.8) are

$$\lim_{h \downarrow 0} \frac{\mathbf{P}(h) - \mathbf{I}}{h} = \mathbf{Q}. \tag{3.10}$$

The matrix $\mathbf{Q}$ is called the *generator*.

The generator for a Markov chain on a state space with four elements is in general

$$\mathbf{Q} = \begin{pmatrix} -q_1 & q_{12} & q_{13} & q_{14} \\ q_{21} & -q_2 & q_{23} & q_{24} \\ q_{31} & q_{32} & -q_3 & q_{34} \\ q_{41} & q_{42} & q_{43} & -q_4 \end{pmatrix}.$$

3.1 An Example: the Poisson process

Let $X = \{X(t) \mid t \geq 0\}$ be a process that has the set of non-negative integers as the state space. One way of defining X as a Poisson process is to assume the following ((1)–(3)).

(1) The increments of the process are independent or

$$P\left(X(t) - X(s), X(u) - X(v)\right) = P\left(X(t) - X(s)\right) \cdot P\left(X(u) - X(v)\right)$$

for $v < u \leq s < t$.

(2) $X(0) = 0$.

(3) $X(t) - X(s) \in \mathrm{Po}(\lambda(t-s))$.

By these assumptions X is a continuous-time Markov chain. This is found by

$$P\left(X(t_n) = j_n | X(t_1) = j_1, X(t_2) = j_2, \ldots, X(t_{n-1}) = j_{n-1}\right) =$$
$$= P\left(X(t_n) - X(t_{n-1}) = j_n - j_{n-1} | X(t_1) = j_1, X(t_2) = j_2, \ldots, X(t_{n-1}) = j_{n-1}\right)$$
$$= P\left(X(t_n) - X(t_{n-1}) = j_n - j_{n-1}\right)$$

by assumptions (1) and (2), and this equals

$$= P\left(X(t_n) - X(t_{n-1}) = j_n - j_{n-1} | X(t_{n-1}) - X(0) = j_{n-1}\right)$$

$$= P\left(X(t_n) = j_n | X(t_{n-1}) = j_{n-1}\right),$$

again by assumption (1). By assumption (3)

$$P_{ij}(t) = P\left(X(t+s) = j | X(s) = i\right) = P\left(X(t+s) - X(s) = j - i\right)$$
$$= e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}.$$

Then from (3.7) we get

$$\lim_{h\downarrow 0}\frac{P_{ii}(h)-1}{h} = \lim_{h\downarrow 0}\frac{e^{-\lambda h}-1}{h} = -\lambda,$$

and from (3.8)

$$\lim_{h\downarrow 0}\frac{P_{ij}(h)}{h} = \lim_{h\downarrow 0}\frac{e^{-\lambda h}\frac{(\lambda h)^{j-i}}{(j-i)!}}{h}$$
$$= \begin{cases} \lambda, & j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

We have the generator

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

A Poisson proces is thus a continuous time Markov chain with $q_i = \lambda$, so that from (3.6) and (3.4) we have

$$P_{ii}(h) = 1 - \lambda h + o(h),$$

or

$$1 - P_{ii}(h) = \lambda h + o(h),$$

so that the total intensity of leaving i is λ. But, by (3), λ is the intensity of one jump upwards for the proces X.

3.2 An Example: The random telegraph

Let $Y \in \mathrm{Be}(1/2)$ and $N = \{N(t) | t \geq 0\}$ be a Poisson process like in the previous example, and let Y be independent of N. Set

$$X(t) = (-1)^{Y+N(t)}.$$

Then $X = \{X(t) \mid t \geq 0\}$ is a continuous-time Markov chain, as esentially follows by the same argument as used in the preceding example. The transition matrix is

$$\mathbf{P}(t) = \begin{pmatrix} \frac{1}{2}\left(1+e^{-2\lambda t}\right) & \frac{1}{2}\left(1-e^{-2\lambda t}\right) \\ \frac{1}{2}\left(1-e^{-2\lambda t}\right) & \frac{1}{2}\left(1+e^{-2\lambda t}\right) \end{pmatrix}.$$

The sample paths of X, the random telegraph, are sequences of -1 and 1, each digit prevailing a random (exponentially distributed time).

4 Forward and backward equations

The meaning of the notion of a generator can be explained as follows. Suppose that $X(0) = i$, and by conditioning $X(t+h)$ on $X(t)$ we get by the Chapman-Kolmogorov equations

$$P_{ij}(t+h) = P\left(X(t+h) = j | X(0) = i\right) =$$

$$= \sum_{k \in \mathcal{X}} P_{ik}(t) P_{kj}(h)$$

$$= P_{ij}(t)\left(1 + q_{jj}h + o(h)\right) + \sum_{k \in \mathcal{X}, k \neq j} P_{ik}(t)\left(q_{kj}h + o(h)\right)$$

from (3.1) and (3.2). Then we get

$$= P_{ij}(t) + h \sum_{k \in \mathcal{X}} P_{ik}(t) q_{kj} + o(h).$$

Thus we get that

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \in \mathcal{X}} P_{ik}(t) q_{kj} + o(h)/h.$$

Hence we have, letting $h \to 0$, and letting $P'_{ij}(t)$ denote the first derivative with respect to t,

$$P'_{ij}(t) = \sum_{k \in \mathcal{X}} P_{ik}(t) q_{kj} = (\mathbf{P}(t)\mathbf{Q})_{ij}.$$

Thus we have derived the following proposition.

Proposition 4.1

$$P'_{ij}(t) = \sum_{k \in \mathcal{X}} P_{ik}(t) q_{kj}, \tag{4.1}$$

or the matrix **forward equation**

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}. \tag{4.2}$$

■

By a similar way, we can prove

Proposition 4.2

$$P'_{ij}(t) = \sum_{k \in \mathcal{X}} q_{ik} P_{kj}(t), \tag{4.3}$$

or the matrix **backward equation**

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t). \tag{4.4}$$

■

Thus we have the system of differential equations

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$$

with the initial conditions

$$P(0) = \mathbf{I}.$$

In courses on differential equations the solution is often written using the exponential of a matrix

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{l=0}^{\infty} \frac{t^l}{l!} \mathbf{Q}^l. \tag{4.5}$$

Hence we see that we can generate $\mathbf{P}(t)$ from knowledge of $\mathbf{Q}$. The representation (4.5) is the unique solution to the forward and backward equations, if the state space is finite.

In many cases the series expansion $e^{\mathbf{Q}t} = \sum_{l=0}^{\infty} \frac{t^l}{l!}\mathbf{Q}^l$ is useless for practical computations, since arbitrary high powers of $\mathbf{Q}$ are needed, see also (Moler and Van Loan 1978). There is, however, a special case of interest in computational evolutionary biology, where the sum can be explicitly evaluated, a fact that leads to important results. We shall develope this in section 7 below.

5 Absolute probabilities

Definition 5.1 *Let $X = \{X(t) \mid t \geq 0\}$ be continuous-time Markov chain. The probability*

$$p_i(t) = P\left(X(t) = i\right)$$

is called the absolute probability for the chain to be in state $i \in \mathcal{X}$ at time t. The vector $\mathbf{p}(t)$ is a row vector whose components are $p_i(t)$. In particular $\mathbf{p}(0)$ is called the initial vector or the initial distribution.

■

The law of total probability gives

$$p_i(t) = P\left(X(t) = i\right) = \sum_{k \in \mathcal{X}} P\left(X(t) = i | X(0) = k\right) P\left(X(0) = k\right) =$$

$$= \sum_{k \in \mathcal{X}} P_{ki}(t) p_k(0),$$

which we write in matrix form as

$$\mathbf{p}(t) = \mathbf{p}(0)\mathbf{P}(t). \tag{5.1}$$

This gives $\mathbf{p}'(t) = \mathbf{p}(0)\mathbf{P}'(t)$. If we multiply the forward equation (4.2) by $\mathbf{p}(0)$ from the left, we get

$$\mathbf{p}'(t) = \mathbf{p}(0)\mathbf{P}'(t) = \mathbf{p}(0)\mathbf{P}(t)\mathbf{Q} = \mathbf{p}(t)\mathbf{Q}. \tag{5.2}$$

If the state space is finite, this is a correct computation, in the case of countable state spaces there are things to be checked. When the equation (5.2) is expressed elementwise, we get

$$p_j'(t) = \sum_{i \in \mathcal{X}} p_i(t) q_{ij} = p_j(t) q_{jj} + \sum_{i \in \mathcal{X}, i \neq j} p_i(t) q_{ij}$$

$$= -q_j p_j(t) + \sum_{i \in \mathcal{X}, i \neq j} p_i(t) q_{ij}.$$

This can be seen as a flow of probabilities. The probability $p_j(t)$ gets an increment that corresponds to the probability that the process is in state i at time t, which is $p_i(t)$ multiplied by the instantaneous transition rate from i to j, q_{ij}. This is summed over all states $i \neq j$. On the other hand $p_j(t)$ is depleted with the probability that the chain is already in the state j multiplied by the instantaneous transition rate to leave the state, i.e., q_j. Inflow minus outfow equals the rate of change $p_j^{'}(t)$.

6 Stationary distribution

6.1 Definition & the global balance equations

Definition 6.1 *The vector* $\pi = (\pi_i)_{i \in \mathcal{X}}$ *is a* stationary distribution *of the chain if*

$$\pi = \pi \mathbf{P}(t) \quad \textit{for all } t \geq 0$$

and $\sum_{i \in \mathcal{X}} \pi_i = 1$ *and* $\pi_i \geq 0$.

■

Thus, (5.1) yields that if $\mathbf{p}(0) = \pi$, then the absolute probabilities are

$$\mathbf{p}(t) = \pi$$

for all $t \geq 0$.

Proposition 6.1

$$\pi = \pi \mathbf{P}(t) \Leftrightarrow \pi \mathbf{Q} = \mathbf{0}. \tag{6.1}$$

Here $\mathbf{0}$ *is matrix of zeros.*

Proof:

$$\pi \mathbf{Q} = \mathbf{0}$$

$$\Leftrightarrow$$

$$\pi \mathbf{Q}^n = \mathbf{0} \quad \text{for all } n \geq 1$$

$$\Leftrightarrow$$

$$\sum_{n=1}^{\infty} \frac{t^n}{n!} \pi \mathbf{Q}^n = \mathbf{0} \quad \text{for all } t \geq 0$$

$$\Leftrightarrow$$

$$\pi \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{Q}^n = \pi \quad \text{for all } t \geq 0$$

$$\Leftrightarrow$$

$$\pi \mathbf{P}(t) = \pi \quad \text{for all } t \geq 0 \ ,$$

as was claimed. ■

If we write $\pi \mathbf{Q} = \mathbf{0}$ elementwise we get

$$\sum_{i \in \mathcal{X}} \pi_i q_{ij} = 0, \quad \text{for every } j,$$

and this we write as

$$\sum_{i \in \mathcal{X}, i \neq j} \pi_i q_{ij} = -\pi_j q_{jj} = \pi_j q_j. \tag{6.2}$$

The left hand side is interpreted as the flow into the state j, since q_{ij} is the instantaneous transition rate from i to j and this is weighted by π_i, which is the probability that the chain is in the state i. These are summed over all states $i \neq j$. In the same way the right hand side is flow out from the state j, since $q_j = \sum_{k \neq j} q_{jk}$ is the total instantaneous transition rate out from that state. A stationary state is reasonably described by inflow being equal to outflow. The system of equations $\pi \mathbf{Q} = \mathbf{0}$ is called the *global balance equations*. The global balance equations or (6.2) will be used in several evolutionary biological contexts in the next lecture.

6.2 Rate of Change

Let the *rate of change*, R, be defined as

$$R \stackrel{\text{def}}{=} \lim_{h \downarrow 0} \frac{P\left(X(t+h) \neq X(t)\right)}{h}. \tag{6.3}$$

We get

$$P\left(X(t+h) \neq X(t)\right) = \sum_{i \in \mathcal{X}} P\left(X(t+h) \neq i \mid X(t) = i\right) P\left(X(t) = i\right) =$$

$$= \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X} j \neq i} P\left(X(t+h) = j \mid X(t) = i\right) P\left(X(t) = i\right).$$

Then, assuming that the state space is finite,

$$R = \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X} j \neq i} \lim_{h \downarrow 0} \frac{P\left(X(t+h) = j \mid X(t) = i\right)}{h} P\left(X(t) = i\right),$$

where from (3.8) we get

$$\lim_{h \downarrow 0} \frac{P\left(X(t+h) = j \mid X(t) = i\right)}{h} = q_{ij}.$$

Thus

$$\begin{aligned} R &= \sum_{i \in \mathcal{X}} \sum_{j \in \mathcal{X} j \neq i} q_{ij} \; P\left(X(t) = i\right) \\ &= - \sum_{i \in \mathcal{X}} q_{ii} P\left(X(t) = i\right), \end{aligned}$$

where we have used (3.4). Assuming stationary state we get the final result.

$$R = - \sum_{i \in \mathcal{X}} q_{ii} \pi_i. \tag{6.4}$$

7 A special generator

We shall next compute the solution to $\pi \mathbf{Q} = \mathbf{0}$ and the rate of change R in (6.4) and $e^{\mathbf{Q}t}$, when the generator is of the form

$$\mathbf{Q} = \begin{pmatrix} -(u-u_1) & u_2 & u_3 & u_4 \\ u_1 & -(u-u_2) & u_3 & u_4 \\ u_1 & u_2 & -(u-u_3) & u_4 \\ u_1 & u_2 & u_3 & -(u-u_4) \end{pmatrix}, \tag{7.1}$$

where

$$u = u_1 + u_2 + u_3 + u_4. \tag{7.2}$$

7.1 Rate of change

Proposition 7.1 *If X is a continuous time Markov chain with the generator $\mathbf{Q}$ in (7.1), then*

$$\pi \mathbf{Q} = \mathbf{0}$$

has the solution

$$\pi = \left(\frac{u_1}{u}, \frac{u_2}{u}, \frac{u_3}{u}, \frac{u_4}{u} \right) \tag{7.3}$$

The proof is left to the reader. ∎

Proposition 7.2 *If X is a continuous time Markov chain with the generator* $\mathbf{Q}$ *in (7.1), then the rate of change is*

$$R = \lim_{h \downarrow 0} \frac{P\left(X(t+h) \neq X(t)\right)}{h} = u\left(1 - \sum_{i \in \mathcal{X}} \pi_i^2\right). \tag{7.4}$$

where u is given in (7.2).

Proof: We have shown in (6.4) that

$$R = -\sum_{i \in \mathcal{X}} q_{ii}\pi_i.$$

If we insert from (7.1) and (7.3) we get

$$-\sum_{i \in \mathcal{X}} q_{ii}\pi_i = \sum_{i \in \mathcal{X}} (u - u_i)\frac{u_i}{u} = u\sum_{i \in \mathcal{X}} (1 - \frac{u_i}{u})\frac{u_i}{u}$$

$$= u\sum_{i \in \mathcal{X}} (1 - \pi_i)\pi_i = u\left(\sum_{i \in \mathcal{X}} \pi_i - \sum_{i \in \mathcal{X}} \pi_i^2\right) = u\left(1 - \sum_{i \in \mathcal{X}} \pi_i^2\right).$$

∎

7.2 The exponential of a generator

Now we find

$$\mathbf{P}(t) = e^{\mathbf{Q}t},$$

when the generator is given in (7.1). For this a couple of smart observations are needed. We introduce the matrix

$$\mathbf{A} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 \\ \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix}.$$

One notes that

$$\mathbf{Q} = -u\left(\mathbf{I} - \mathbf{A}\right). \tag{7.5}$$

The interesting thing about $\mathbf{A}$ is, that it is idempotent, i.e.,

$$\mathbf{A}^n = \mathbf{A}, \qquad \text{for } n \geq 1.$$

This is easily verified by a computation

$$\mathbf{A}^2 = \begin{pmatrix} \pi_1(\pi_1+\pi_2+\pi_3+\pi_4) & \ldots\ldots & \pi_4(\pi_1+\pi_2+\pi_3+\pi_4) \\ \pi_1(\pi_1+\pi_2+\pi_3+\pi_4) & \ldots\ldots & \pi_4(\pi_1+\pi_2+\pi_3+\pi_4) \\ \pi_1(\pi_1+\pi_2+\pi_3+\pi_4) & \ldots\ldots & \pi_4(\pi_1+\pi_2+\pi_3+\pi_4) \\ \pi_1(\pi_1+\pi_2+\pi_3+\pi_4) & \ldots\ldots & \pi_4(\pi_1+\pi_2+\pi_3+\pi_4) \end{pmatrix}$$

$$= \mathbf{A}.$$

Thus $\mathbf{A}^n = \mathbf{A}$ for all $n \geq 1$.

Next we recall that

$$e^{-ut\mathbf{I}} = \sum_{l=0}^{\infty} \frac{(-ut)^l}{l!}\mathbf{I}^l = \mathbf{I}\sum_{l=0}^{\infty} \frac{(-ut)^l}{l!} = e^{-ut}\mathbf{I}.$$

Then we have

$$\begin{aligned} e^{\mathbf{Q}t} &= e^{-ut(\mathbf{I}-\mathbf{A})} = e^{-ut}\mathbf{I}e^{ut\mathbf{A}} \\ &= e^{-ut}\mathbf{I}\sum_{l=0}^{\infty}\frac{(ut)^l}{l!}\mathbf{A}^l = e^{-ut}\mathbf{I}\left[\mathbf{I} + \sum_{l=1}^{\infty}\frac{(ut)^l}{l!}\mathbf{A}^l\right] \\ &= e^{-ut}\mathbf{I}\left[\mathbf{I} + \mathbf{A}\sum_{l=1}^{\infty}\frac{(ut)^l}{l!}\right] = e^{-ut}\mathbf{I}\left[\mathbf{I} + \mathbf{A}(e^{-ut}-1)\right] \\ &= e^{-ut}\mathbf{I} + \mathbf{A}\left(1-e^{-ut}\right). \end{aligned}$$

To summarize

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = e^{-ut}\mathbf{I} + \mathbf{A}\left(1-e^{-ut}\right), \qquad \text{for } \mathbf{Q} \text{ in (7.1).} \tag{7.6}$$

If we write this elementwise, we get

$$P_{ij}(t) = e^{-ut}\delta_{i,j} + \left(1-e^{-ut}\right)\pi_j, \tag{7.7}$$

where $\delta_{i,j}$ is the Kronecker delta defined by

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \tag{7.8}$$

7.3 Separation of Species

We are still assuming $\mathbf{Q}$ in (7.1), and make the additional assumption of *reversibility*. We assume namely first that

$$\pi_i P_{ij}(t) = \pi_j P_{ji}(t) \quad \text{for all } t,\ i, j\ \mathcal{X}. \tag{7.9}$$

This implies by (3.8) even that

$$\pi_i q_{ij} = \pi_j q_{ji}.$$

We see immediately that this is satisfied for $\mathbf{Q}$ in (7.1).

Let us now suppose that we have two continuous-time Markov chains X and Y, with the same generator $\mathbf{Q}$ in (7.1), assuming reversibility, and such that

$$X(0) = Y(0).$$

Then we have

Proposition 7.3 *Assume two continuous-time Markov chains X and Y, with the same generator $\mathbf{Q}$ in (7.1), assuming reversibility, and such that*

$$X(0) = Y(0) \in \pi,$$

but evolving independently thereafter. Then

$$P\left(X(t) = i, Y(t) = j\right) = \pi_i P_{ij}(2t) =$$

$$= \begin{cases} \pi_i \left(1 - e^{-2ut}\right) \pi_j & i \neq j, \\ \pi_i e^{-2ut} + \pi_i \left(1 - e^{-2ut}\right) \pi_j & i = j. \end{cases} \tag{7.10}$$

Proof:

$$P\left(X(t) = i, Y(t) = j\right) = \sum_{k \in \mathcal{X}} P\left(X(t) = i, Y(t) = j \mid X(0) = Y(0) = k\right) P\left(X(0) = Y(0) = k\right)$$

$$= \sum_{k \in \mathcal{X}} P\left(X(t) = i, Y(t) = j \mid X(0) = Y(0) = k\right) \pi_k$$

$$= \sum_{k \in \mathcal{X}} P\left(X(t) = i \mid X(0) = k\right) P\left(Y(t) = j \mid Y(0) = k\right) \pi_k$$

$$= \sum_{k \in \mathcal{X}} P_{ki}(t) P_{kj}(t) \pi_k =$$

$$= \pi_i \sum_{k \in \mathcal{X}} P_{ik}(t) P_{kj}(t)$$

where we used reversibility (7.9), and the right hand side is

$$= \pi_i P_{ij}(2t),$$

where we invoked the Chapman-Kolmogorov equations (2.7). Then we insert from (7.7) to obtain the rightmost expression in (7.10). ■

7.4 Fraction of divergence

Next we compute

$$P\left(X(t) \neq Y(t)\right).$$

From the preceding proposition we get

$$P\left(X(t) \neq Y(t)\right) = \sum_{i \neq j} P\left(X(t) = i, Y(t) = j\right)$$

$$= \sum_{i \neq j} \pi_i P_{ij}(2t) = \sum_{i \neq j} \pi_i \left(1 - e^{-2ut}\right) \pi_j,$$

in view of (7.10). The right hand side equals

$$= \left(1 - e^{-2ut}\right) \sum_{i \neq j} \pi_i \pi_j.$$

Here the sum $\sum_{i \neq j} \pi_i \pi_j$ is actually a double sum

$$\sum_{i \neq j} \pi_i \pi_j = \sum_{i \in \mathcal{X}} \pi_i \sum_{j \in \mathcal{X}, j \neq i} \pi_j =$$

$$= \sum_{i \in \mathcal{X}} \pi_i (\pi_1 + \pi_2 + \ldots \pi_{i-1} + \pi_{i+1} + \ldots) =$$

$$= \sum_{i \in \mathcal{X}} \pi_i (1 - \pi_i) = \sum_{i \in \mathcal{X}} \pi_i - \sum_{i \in \mathcal{X}} \pi_i^2 = 1 - \sum_{i \in \mathcal{X}} \pi_i^2.$$

Now we recall the rate of change from (7.4), and get

$$P\left(X(t) \neq Y(t)\right) = \frac{R}{u} \left(1 - e^{-2ut}\right).$$

This will lead to a famous formula of evolutionary biology, *Jukes-Cantor formula*, later on.

8 Exponential holding times

Assume that $X(t) = i$. Let

$$T = \inf\{s \geq 0 | X(t+s) \neq i\}.$$

Then T is the further time the continuous time chain remains in the state i. The time T is called the *holding time*.

Proposition 8.1 $T|X(t) = i \in \text{Exp}\,(q_j)$.

Proof: We start by

$$P\,(T \geq t+h) = P\,(T \geq t)\,P\,(T \geq t+h|T \geq t)\,.$$

The continuous-time Markov chain lacks memory and is time-homogeneous. Hence we get that

$$P\,(T \geq t+h) = P\,(T \geq t)\,P\,(T \geq h)\,.$$

But the probability $P\,(T \geq h)$ is the probability that the continuous-time Markov chain has not moved from i before the time h. From the assumption (3.2) we obtain

$$P\,(T \geq h) = P_{ii}(h) = 1 + q_{ii}h + o(h) = 1 - q_i h + o(h).$$

Hence

$$P\,(T \geq t+h) = P\,(T \geq t)\,(1 - q_i h) + o(h).$$

This is rearranged as

$$\frac{P\,(T \geq t+h) - P\,(T \geq t)}{h} = -q_i P\,(T \geq t) + \frac{o(h)}{h}.$$

Letting $h \to 0$ this yields

$$\frac{d}{dt} P\,(T \geq t) = -q_i P\,(T \geq t)\,.$$

This differential equation has the general solution

$$P\,(T \geq t) = C e^{-q_i t},$$

for some constant C. For $t = 0$ we get $C = 1$. Thus

$$P(T \geq t) = e^{-q_i t},$$

or

$$P(T \leq t) = 1 - e^{-q_i t},$$

which 'verifies' the claim as asserted. ∎

The mean time spent at i is thus $\frac{1}{q_j}$. Assume that $X(0) = i$. Let

$$T_1 = \inf\{s \geq 0 | X(s) \neq i\}.$$

Proposition 8.2

$$P(X(T_1) = j | X(0) = i) = \frac{q_{ij}}{q_i}, \quad j \neq i.$$

Sketch of Proof: Let us define for $i \neq j$,

$$r_{ij}(h) = P(X(h) = j | X(0) = i, X(h) \neq i),$$

Next, if $0 \leq T_1 \leq h$,

$$P(X(T_1) = j | X(0) = i) = P(\text{ } X \text{ jumps to } j \mid X \text{ jumps only once in } [0, h] \text{ })$$
$$\approx r_{ij}(h)$$

for small h. But

$$r_{ij}(h) = P(X(h) = j | X(0) = i, X(h) \neq i) = \frac{P(X(h) = j, X(0) = i, X(h) \neq i)}{P(X(0) = i, X(h) \neq i)}$$
$$= \frac{P(X(h) = j, X(0) = i)}{P(X(0) = i, X(h) \neq i)}$$
$$= \frac{P(X(h) = j | X(0) = i)}{P(X(h) \neq i | X(0) = i)}$$
$$= \frac{P_{ij}(h)}{1 - P_{ii}(h)} = \frac{P_{ij}(h)/h}{(1 - P_{ii}(h))/h}.$$

For small h

$$\frac{P_{ij}(h)/h}{(1 - P_{ii}(h))/h} \approx \frac{q_{ij}}{-q_{ii}} = \frac{q_{ij}}{q_i},$$

as was claimed. ∎

9 The imbedded Markov chain

From the preceding, if $X(t) = i$, it remains there for an exponentially distributed time T and jumps then to another state with the probability $\frac{q_{ij}}{q_i}$. Let us set

$$\tilde{X}_0 = X(0), T_0 = 0.$$

Take $n \geq 1$. Suppose that $\tilde{X}_{n-1}$ has been defined. Let $U_n \in \mathrm{Exp}(q_i)$. Let

$$T_n = T_{n-1} + U_n$$

and

$$\tilde{X}_n = X\left(T_n\right).$$

By lack of memory and time homogeneity $\{\tilde{X}_n\}_{n=0}^{\infty}$ is a Markov chain that jumps from i to j with probability $\frac{q_{ij}}{q_i}$, if $i \neq j$ and jumps from i to i with probability 0. This is called the *imbedded Markov chain*.

10 References

G.R. Grimmett and D.R. Stirzaker (1994): *Probability and Random Processes 2nd Ed.*. Oxford Science Publications. Clarendon Press, Oxford.
C.B. Moler and C.F. Van Loan (1978): Nineteen Dubious Ways to Compute the Exponential of a Matrix. *SIAM Review*, 20, pp. 801−836.
J.R. Norris (1998): *Markov Chains.* Cambridge University Press.
M. Waterman (1995): *Introduction to Computational Biology. Maps, Sequences and Genomes.* Chapman and Hall, London, et.al.,