2.1.4 Fundamental Matrices

2.1.4.1 Irreducible Markov chains

Definition. Let $\{X_n\}$ be an irreducible aperiodic Markov chain on a finite state space with transition probability matrix $\mathbf{P}$. Then

$$\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{1}\boldsymbol{\pi}^T)^{-1}$$

is called the **fundamental matrix of the irreducible Markov chain**.

Proposition 2.6. *The fundamental matrix* $\mathbf{Z}$ *is well defined and* $Z = \mathbf{I} + \sum_{n=1}^{\infty}(\mathbf{P}^n - \mathbf{1}\boldsymbol{\pi}^T)$.

Proof. First, notice that $\mathbf{1}\boldsymbol{\pi}^T\mathbf{P} = \mathbf{1}\boldsymbol{\pi}^T$, $\mathbf{1}\boldsymbol{\pi}^T\mathbf{P} = \mathbf{1}\boldsymbol{\pi}^T$, $(\mathbf{1}\boldsymbol{\pi}^T)^n = \mathbf{1}\boldsymbol{\pi}^T$, and $\mathbf{P}(\mathbf{1}\boldsymbol{\pi}^T)^n = (\mathbf{1}\boldsymbol{\pi}^T)^n\mathbf{P} = \mathbf{1}\boldsymbol{\pi}^T$. Then,

$$(\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T)^n = \sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}\mathbf{P}^k(\mathbf{1}\boldsymbol{\pi}^T)^{n-k} = \mathbf{P}^n + \underbrace{\left[\sum_{k=0}^{n}\binom{n}{k}(-1)^{n-k}\right]}_{=-1}\mathbf{1}\boldsymbol{\pi}^T = \mathbf{P}^n - \mathbf{1}\boldsymbol{\pi}^T.$$

Therefore,

$$[\mathbf{I} - (\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T)][\mathbf{I} + (\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T) + \cdots + (\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T)^{n-1}] = \mathbf{I} - \mathbf{P}^n + \mathbf{1}\boldsymbol{\pi}^T$$

Sending n to ∞, we arrive at

$$[\mathbf{I} - (\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T)]\left[\mathbf{I} + \sum_{n=1}^{\infty}(\mathbf{P} - \mathbf{1}\boldsymbol{\pi}^T)^n\right] = \mathbf{I},$$

so

$$Z = (\mathbf{I} - \mathbf{P} + \mathbf{1}\boldsymbol{\pi}^T)^{-1} = \mathbf{I} + \sum_{n=1}^{\infty}(\mathbf{P}^n - \mathbf{1}\boldsymbol{\pi}^T)$$

exists. □

Note 2.8. If we define $m_{ij} = E_i(T_j)$, then

$$m_{ij} = \frac{z_{jj} - z_{ij}}{\pi_j}.$$

TO DO: Add a remark about periodicity and basic limit theorem.

2.2 Continuous-time Markov chains

2.2.1 Motivation

First let us ask ourselves why we need continuous-time Markov models. These models are preferable to their discrete-time counterparts if there is no natural unit of time that can be used as a length of the time step. This happens frequently during statistical modeling of unevenly spaced observations. Continuous-time models naturally handle unevenly spaced observations without discretization via an evenly spaced grid, which leads to introduction of missing data at some some of the grid points. In addition, in applications where deterministic modeling via differential equations is common, precise connections between deterministic and stochastic models can be established.

2.2.2 Definition and properties

Definition. $\{X_t\}$, $t \geq 0$ is a **continuous-time Markov chain** if for any $0 \leq s_0 < s_1 < \cdots s_n < s$ and possible states $i_0, \ldots, i_n, i, j \in \Omega$

$$\Pr\left(X_{t+s} = j \,|\, X_s = i, X_{s_n} = i_n, \ldots, X_{s_0} = i_0\right) = \Pr\left(X_{t+s} = j \,|\, X_s = i\right).$$

If $\Pr(X_{t+s} = j \,|\, X_s = i) = \Pr(X_t = j \,|\, X_0 = i)$ for all $s \geq 0$, then $\{X_t\}$ is called **homogeneous**. In such a case, $\Pr(X_t = j \,|\, X_0 = i) = p_{ij}(t)$ are called **finite-time transition probabilities** and $\mathbf{P}(t) = \{p_{ij}(t)\}$ is called **transition probability matrix** or transition semigroup.

Note 2.9. A semigroup is a set $\mathscr{S}$ with an algebraic operation "$\times$" such that

1. $\forall a, b \in \mathscr{S}$, $a \times b \in S$ — this property is called closure
2. $\forall a, b, c \in \mathscr{S}$, $(a \times b) \times c = a \times (b \times c)$ — this property is called associativity

Associativity holds for all matrices, not just for transition probability matrices. The next result shows that the set of all possible transition probability matrices is closed under matrix multiplication.

Chapman-Kolmogorov equation

$p_{ij}(t+s) = \sum_{k \in \Omega} p_{ik}(s) p_{kj}(t)$ or $\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$.

Proof.

$$\begin{aligned}
&p_{ij}(t+s) = \Pr(X_{t+s} = j \,|\, X_0 = i) = \sum_{k \in \Omega} \Pr(X_{t+s} = j, X_s = k \,|\, X_0 = i) \\
&= \sum_{k \in \Omega} \Pr(X_{t+s} = j \,|\, X_s = k, X_0 = i) \Pr(X_s = k \,|\, X_0 = i) = [\text{Markov property + homogeneity}] \\
&= \sum_{k \in \Omega} p_{ik}(s) p_{kj}(t).
\end{aligned}$$

□

Note 2.10. $\mathbf{P}(0) = \mathbf{I}$, because $p_{ij}(0) = \Pr(X_0 = j \,|\, X_0 = i) = 1_{\{i=j\}}$.

Theorem 2.5. *Let $\mathbf{P}(t)$ be a transition probability matrix with $\lim_{t \to 0^+} \mathbf{P}(t) = \mathbf{P}(0) = \mathbf{I}$, then $\forall i \in \Omega$, there exists*

$$\lambda_i \overset{def}{=} \lim_{h \to 0^+} \frac{1 - p_{ii}(h)}{h} \in [0, \infty]$$

and $\forall i \neq j \in \Omega$, there exists

$$\lambda_{ij} \overset{def}{=} \lim_{h \to 0^+} \frac{p_{ij}(h)}{h} \in [0, \infty].$$

Definition. $\mathbf{\Lambda} = \{\lambda_{ij}\}_{i,j \in \Omega}$, with $\lambda_{ii} \overset{\text{def}}{=} -\lambda_i$, is called the **infinitesimal generator** of the semigroup.

1. $\{X_t\}$ is called **stable** if $\lambda_i < \infty \; \forall i \in \Omega$.
2. $\{X_t\}$ is called **conservative** if $\lambda_i = \sum_{j \neq i} \lambda_{ij} \; \forall i \in \Omega$.

We will work with stable and conservative continuous-time Markov chains. For such chains, we can write

$$\begin{aligned}
p_{ii}(t) &= 1 - \lambda_i t + o(t), \\
p_{ij}(t) &= \lambda_{ij} t + o(t).
\end{aligned}$$

Using Chapman-Kolmogorov equation, we get

$$\mathbf{P}'(t) = \lim_{h\to 0}\frac{\mathbf{P}(t+h)-\mathbf{P}(t)}{h} = \lim_{h\to 0}\frac{\mathbf{P}(t)\mathbf{P}(h)-\mathbf{P}(t)}{h} = \lim_{h\to 0}\frac{\mathbf{P}(t)[\mathbf{P}(h)-\mathbf{I}]}{h} = \mathbf{P}(t)\mathbf{\Lambda}.$$

This equation is called Kolmogorov forward equation. The backward equation, $\mathbf{P}'(t) = \mathbf{\Lambda}\mathbf{P}(t)$, is obtained by exchanging the order of multiplication on the right-hand side. These two differential equations together with the initial condition $\mathbf{P}(0) = \mathbf{I}$ yield a unique solution:

$$\mathbf{P}(t) = e^{\mathbf{\Lambda}t} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty}\frac{(\mathbf{\Lambda}t)^k}{k!}.$$

The fact that this matrix exponential is indeed a solution of the forward and backward equations can be verified by differentiation. Uniqueness comes from standard theory of differential equations.

Note 2.11. The definition of a conservative continuous-time Markov chain automatically translates into $\mathbf{P}(t)\mathbf{1} = \mathbf{1}$, because

$$\mathbf{\Lambda 1} = \mathbf{0} \Rightarrow \mathbf{P}'(t)\mathbf{1} = \mathbf{P}(t)\mathbf{\Lambda 1} = \mathbf{P}(t)\mathbf{0} = \mathbf{0} \Rightarrow [\mathbf{P}(t)\mathbf{1}]' = \mathbf{0} \Rightarrow \mathbf{P}(t)\mathbf{1} = \mathbf{c}.$$

But $\mathbf{c} = \mathbf{P}(0)\mathbf{1} = \mathbf{I1} = \mathbf{1}$, which proves that $\mathbf{P}(t)\mathbf{1} = \mathbf{1}$.

Example: *two-state chain* Let $\{X_t\}$ be a continuous-time Markov chain on $\Sigma = \{1,2\}$ with infinitesimal generator

$$\mathbf{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}.$$

The backward equation yields

$$\begin{pmatrix} p'_{11}(t) & p'_{12}(t) \\ p'_{21}(t) & p'_{22}(t) \end{pmatrix} = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}\begin{pmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{pmatrix}.$$

Since $\mathbf{P}(t)\mathbf{1} = \mathbf{1}$ we need to solve only the following system of ordinary differential equations:

$$\begin{aligned} p'_{11}(t) &= -\lambda_1[p_{11}(t) - p_{21}(t)], \\ p'_{21}(t) &= \lambda_2[p_{11}(t) - p_{21}(t)]. \end{aligned}$$

Subtracting the second equation from the first one, we get

$$p'_{11}(t) - p'_{21}(t) = -(\lambda_1+\lambda_2)[p_{11}(t) - p_{21}(t)] \Rightarrow p_{11}(t) - p_{21}(t) = e^{-(\lambda_1+\lambda_2)t},$$

because $p_{11}(0) - p_{21}(0) = 1 - 0 = 1$.

$$p'_{21}(t) = \lambda_2[p_{11}(t) - p_{21}(t)] = \lambda_2 e^{-(\lambda_1+\lambda_2)t} \Rightarrow p_{21}(t) = -\frac{\lambda_2}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)t} + C.$$

The normalizing constant C is found with the help of the initial condition:

$$p_{21}(0) = -\frac{\lambda_2}{\lambda_2+\lambda_1} + C = 0 \Rightarrow C = \frac{\lambda_2}{\lambda_1+\lambda_2}.$$

So far, we proved that

$$p_{21}(t) = \frac{\lambda_2}{\lambda_1+\lambda_2} - \frac{\lambda_2}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)t}.$$

$$p_{11}(t) = e^{-(\lambda_1+\lambda_2)t} + p_{21}(t) = \frac{\lambda_2}{\lambda_1+\lambda_2} + \frac{\lambda_1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)t}.$$

Therefore,

$$\mathbf{P}(t) = \frac{1}{\lambda_1+\lambda_2}\begin{pmatrix} \lambda_2+\lambda_1 e^{-(\lambda_1+\lambda_2)t} & \lambda_1-\lambda_1 e^{-(\lambda_1+\lambda_2)t} \\ \lambda_2-\lambda_2 e^{-(\lambda_1+\lambda_2)t} & \lambda_1+\lambda_2 e^{-(\lambda_1+\lambda_2)t} \end{pmatrix}.$$

This example shows that sometimes we can derive explicit algebraic formulas for transition probabilities as functions of infinitesimal transition rates. This is useful in statistical applications, as we will see later.

2.2.3 Simulation and discretization of continuous-time Markov chains

We can construct the CTMC process $(X_u, 0 \le u \le t)$ with initial distribution $\boldsymbol{\mu}$ and intensity matrix $\boldsymbol{\Lambda}$. The procedure is illustrated in Figure 2.6

Algorithm 1 Simulate CTMC path (Gillespie algorithm)

Choose initial state i_{new} by generating a discrete random variable with probability mass function specified by $\boldsymbol{\mu}$. Let $X_0 = i_{\text{new}}$
Set PROCEED=TRUE
Set the state vector $\mathbf{x} = (i_{\text{new}})$
Set the waiting time vector $\boldsymbol{\tau} = (0)$
Set $\tau = 0$ and $\tau_{\text{cur}} = 0$
while PROCEED **do**
    if $\lambda_{i_{\text{new}}} = 0$ or $\tau_{\text{cur}} \ge t$ **then**
        Append i_{new} to the state vector $\mathbf{x}$
        Append t to the waiting time vector $\boldsymbol{\tau}$
        Set PROCEED=FALSE
    else
        Draw an exponentially distributed waiting time $\tau \sim \text{Exp}(\lambda_{i_{\text{new}}})$
        Set $\tau_{\text{cur}} = \tau_{\text{cur}} + \tau$
        Append τ_{cur} to the waiting time vector $\boldsymbol{\tau}$
        Set i_{new} to a new value by generating a discrete random variable with probability mass function defined by probabilities $r_{jk} = \lambda_{jk}/\lambda_j$, for $k \in \Omega$
        Append i_{new} to the state vector $\mathbf{x}$
return waiting time vector $\boldsymbol{\tau}$ and state vector $\mathbf{x}$.

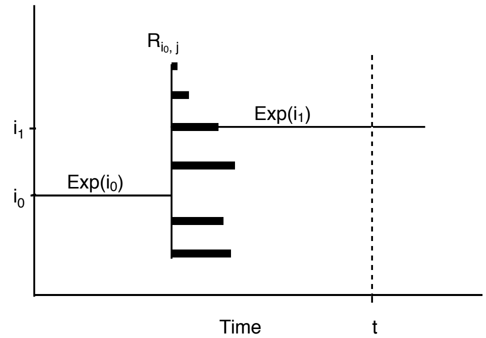

FIGURE 2.6: Construction of a CTMC.

Connection to the transition probability postulates:

$$\Pr(\text{no jumps in } [0,t) \,|\, X_0 = i) = \Pr(T_i > t) = e^{-\lambda_i t} = \sum_{n=0}^{\infty} \frac{(-\lambda_i t)^n}{n!} = 1 - \lambda_i t + o(t),$$

$$\Pr(\text{one jump to } j \neq i \text{ in } [0,t) \,|\, X_0 = i) = \Pr(T_i < t)\, \frac{\lambda_{ij}}{\lambda_i} \Pr(\text{no jump from } j \,|\, T_i < t)$$

$$= (1 - e^{-\lambda_i t}) \frac{\lambda_{ij}}{\lambda_i} \int_0^t e^{\lambda_j (t-\tau)} \frac{\lambda_i}{1 - e^{-\lambda_i t}} e^{-\lambda_i \tau} d\tau = \lambda_{ij} e^{-\lambda_j t} \frac{1}{\lambda_j - \lambda_i} \left[e^{(\lambda_j - \lambda_i)t} - 1 \right]$$

$$= \lambda_{ij} [1 - \lambda_j t + o(t)] \frac{1}{\lambda_j - \lambda_i} [(\lambda_j - \lambda_i)t + o(t)] = \lambda_{ij} t + o(t),$$

$$\Pr(\text{more than one jump in } [0,t) \,|\, X_0 = i) = 1 - \Pr(\text{no jumps in } [0,t) \,|\, X_0 = i)$$

$$- \sum_{j \neq i} \Pr(\text{one jump to } j \neq i \text{ in} [0,t) \,|\, X_0 = i)$$

$$= 1 - [1 - \lambda_i t + o(t)] - \sum_{j \neq i} [\lambda_{ij} t + o(t)] = o(t).$$

Therefore,

$$p_{ii}(t) = \Pr(T_i > t) + o(t) = e^{-\lambda_i t} + o(t) = 1 - \lambda_i t + o(t),$$
$$p_{ij}(t) = \Pr(\text{one jump to } j \neq i \text{ in } [0,t) \,|\, X_0 = i) + o(t) = \lambda_{ij} t + o(t).$$

Skeleton discretizaton:

We simply ignore times between transitions and define a discrete-time Markov chain with transition probabilities

$$r_{ij} = \begin{cases} \frac{\lambda_{ij}}{\lambda_i} & \text{if } i \neq j, \\ 0 & \text{if } i = j, \end{cases}$$

where $\mathbf{\Lambda} = \{\lambda_{ij}\}$ is the generator of the continuous-time Markov chain $\{X_t\}$.

Discretization via uniformization:

Choose $\mu \geq \max_i \{\lambda_i\}$. Define a transition probability matrix

$$\mathbf{Q} = \mathbf{I} + \frac{1}{\mu} \mathbf{\Lambda}.$$

Let Z_n be a DTMC with this transition probability matrix. Let's first confirm that $\mathbf{Q}$ is indeed a stochastic matrix. First,

$$\mathbf{Q1} = \mathbf{I1} + \frac{1}{\mu}\mathbf{\Lambda 1} = \mathbf{1} + \mathbf{0} = \mathbf{1}.$$

Next, $q_{ij} = \lambda_{ij}/\mu \geq 0$ for $i \neq j$ and

$$q_{ii} = 1 - \frac{\lambda_i}{\mu} \geq 1 - \frac{\lambda_i}{\max_i\{\lambda_i\}} \geq 1 - \frac{\lambda_i}{\lambda_i} = 0.$$

So Z_n is well defined, but how does it relate to the original continuous-time process X_t?

Proposition 2.7. *Let N_t be a with rate μ. We assume that this process is independent of Z_n. Then Z_{N_t} is continuous-time Markov chain with the generator $\mathbf{\Lambda}$.*

Proof. We want to prove that $\Pr(Z_{N_t} = j \,|\, Z_0 = i) = \Pr(X_t = j \,|\, X_0 = i)$. Let $\mathbf{P}(t) = e^{\mathbf{\Lambda} t}$ be the finite-time transition probability matrix for X_t and let $\tilde{\mathbf{P}}(t) = \{\tilde{p}_{ij}(t)\}$ be the corresponding matrix for Z_{N_t} with $\tilde{p}_{ij}(t) = \Pr(Z_{N_t} = j \,|\, Z_0 = i)$. Both transition probability matrices satisfy the same initial condition: $\mathbf{P}(0) = \tilde{\mathbf{P}}(0) = \mathbf{I}$. We would like to show that $\tilde{\mathbf{P}}'(t) = \mathbf{\Lambda}\tilde{\mathbf{P}}(t)$. Since $\mathbf{P}(t)$ satisfies the same differential equation, $\mathbf{P}(t)$ and $\tilde{\mathbf{P}}(t)$ must be equal. To prove that $\tilde{\mathbf{P}}'(t) = \mathbf{\Lambda}\tilde{\mathbf{P}}(t)$ we write

$$\tilde{p}_{ij}(t) = \Pr(Z_{N_t} = j \,|\, Z_0 = i) = \sum_{n=0}^{\infty} \Pr(Z_n = j \,|\, Z_0 = i, N_t = n) \Pr(N_t = n) = \sum_{n=0}^{\infty} q_{ij}^{(n)} e^{-\mu t} \frac{(\mu t)^n}{n!}.$$

The proof is finished by writing the above equation in a matrix form. □

2.2.4 Stationarity and reversibility

Definition. The probability vector $\boldsymbol{\pi}$ is called a **stationary distribution** of the continuous Markov chain X_t if $\boldsymbol{\pi}^T \mathbf{P}(t) = \boldsymbol{\pi}^T$ for all $t \geq 0$.

Proposition 2.8. *$\boldsymbol{\pi}$ is a stationary distribution $\Leftrightarrow$ $\boldsymbol{\pi}^T \mathbf{\Lambda} = \mathbf{0}^T$.*

Proof. Suppose $\boldsymbol{\pi}$ is a stationary distribution. Then the forward equation yields

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{\Lambda} \Rightarrow \boldsymbol{\pi}^T \mathbf{\Lambda} = \boldsymbol{\pi}^T \mathbf{P}(t)\mathbf{\Lambda} = \boldsymbol{\pi}^T \mathbf{P}'(t) = \left[\boldsymbol{\pi}^T \mathbf{P}(t)\right]' = \left[\boldsymbol{\pi}^T\right]' = \mathbf{0}^T.$$

Now, suppose $\boldsymbol{\pi}^T \mathbf{\Lambda} = \mathbf{0}^T$. Then

$$\boldsymbol{\pi}^T \mathbf{P}(t) = \boldsymbol{\pi}^T \sum_{k=0}^{\infty} \frac{(\mathbf{\Lambda} t)^k}{k!} = \boldsymbol{\pi}^T \mathbf{I} + \sum_{k=1}^{\infty} \frac{\boldsymbol{\pi}^T \mathbf{\Lambda} \mathbf{\Lambda}^{k-1} t^k}{k!} = \boldsymbol{\pi}^T \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{0}^T \mathbf{\Lambda}^{k-1} t^k}{k!} = \boldsymbol{\pi}^T.$$

□

Definition. $\{X_t\}$ is called **irreducible** if all states communicate: for all $i, j \in \Omega$ there exists $t > 0$ such that $p_{ij}(t) > 0$.

As before, an irreducible CTMC on a finite state space has a unique stationary distribution. One can also formulate continuous-time ergodic theorems. However, such limiting result does not find extensive use in statistics.

Definition. The infinitesimal generator $\mathbf{\Lambda}$ is said to satisfy **detailed balance** if there exists a probability vector $\boldsymbol{\pi}$ such that $\lambda_{ij}\pi_i = \lambda_{ji}\pi_j$.

As in the discrete-time Markov chains, if $\boldsymbol{\pi}$ and $\mathbf{\Lambda}$ satisfy detailed balance, then $\boldsymbol{\pi}$ is a stationary distribution. You should be able to prove it the same way we have proved the analogous statement for DTMCs.

Example: *Two-state CTMC continued* Recall that the generator and finite-time transition probability matrices for this model are

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix} \text{ and } \mathbf{P}(t) = \frac{1}{\lambda_1 + \lambda_2} \begin{pmatrix} \lambda_2 + \lambda_1 e^{-(\lambda_1+\lambda_2)t} & \lambda_1 - \lambda_1 e^{-(\lambda_1+\lambda_2)t} \\ \lambda_2 - \lambda_2 e^{-(\lambda_1+\lambda_2)t} & \lambda_1 + \lambda_2 e^{-(\lambda_1+\lambda_2)t} \end{pmatrix}.$$

We have already noticed that if we send t to infinity, $\mathbf{P}(t)$ approaches a matrix with rows equal to

$$\boldsymbol{\pi}^T = \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}, \frac{\lambda_1}{\lambda_1 + \lambda_2} \right).$$

Checking detailed balance:

$$\lambda_1 \times \frac{\lambda_2}{\lambda_1 + \lambda_2} = \lambda_2 \times \frac{\lambda_1}{\lambda_1 + \lambda_2},$$

we confirm that $\boldsymbol{\pi}^T$ is indeed a stationary distribution of this CTMC. When $\lambda_1 > 0$ and $\lambda_2 > 0$, the chain is irreducible and the stationary distribution is unique.

Example: *Oxygen attachment to hemoglobin* Hemoglobin has 4 possible oxygen binding sites. Let $X_t \in \{0, 1, 2, 3, 4\}$ be the number of occupied O_2 binding sites in a given hemoglobin molecule. Let s_0 be the concentration of O_2. We model oxygen attachment as a CTMC with generator

$$\boldsymbol{\Lambda} = \begin{pmatrix} - & s_o k_{+1} & 0 & 0 & 0 \\ k_{-1} & - & s_0 k_{+2} & 0 & 0 \\ 0 & k_{-2} & - & s_0 k_{+3} & 0 \\ 0 & 0 & k_{-3} & - & s_0 k_{+4} \\ 0 & 0 & 0 & k_{-4} & - \end{pmatrix}$$

This irreducible chain lives on a finite state space, so we know that there exists a stationary distribution $\boldsymbol{\pi}^T = (\pi_0, \ldots, \pi_4)$ and it is unique. Let's try detailed balance first:

$$\begin{aligned} \pi_0 s_o k_{+1} &= \pi_1 k_{-1} \Rightarrow \pi_1 = \pi_0 s_0 \frac{k_{+1}}{k_{-1}} \\ \pi_1 s_o k_{+2} &= \pi_2 k_{-2} \Rightarrow \pi_2 = \pi_1 s_0 \frac{k_{+2}}{k_{-2}} = \pi_0 s_0^2 \frac{k_{+1} k_{+2}}{k_{-1} k_{-2}} \\ &\vdots \\ \pi_3 s_o k_{+4} &= \pi_4 k_{-4} \Rightarrow \pi_4 = \pi_0 s_0^4 \prod_{j=1}^{4} \frac{k_{+j}}{k_{-j}} \end{aligned}$$

Using $\sum_{i=0}^4 \pi_i = 1$, we get

$$\pi_0 \left[1 + \sum_{i=1}^{4} s_0^i \prod_{j=1}^{i} \frac{k_{+j}}{k_{-j}} \right] = 1 \Rightarrow \pi_0 = \frac{1}{1 + \sum_{i=1}^4 s_0^i \prod_{j=1}^i \frac{k_{+j}}{k_{-j}}}.$$

So we have found the stationary distribution and proved that $\{X_t\}$ is reversible on the way.

Note 2.12. We need 4 different binding rates and 4 different association rates to account for positive cooperativity of O_2 binding.

2.2.5 Matrix exponentiation

Here, we are going to discuss two popular methods for exponentiating infinitesimal generators. See (Moler and Loan, 2003) for a good review of matrix exponentiation methods.

Method 1: series truncation

Use the definition of matrix exponential and set

$$e^{\mathbf{\Lambda}t} \approx \sum_{n=0}^{k} \frac{(\mathbf{\Lambda}t)^n}{n!}$$

for some fixed k. This naive method could very inaccurate for some choices of $\mathbf{\Lambda}$ and t, because $\mathbf{\Lambda}$ has both positive and negative entries. Instead, we can use uniformization:

$$e^{\mathbf{\Lambda}t} \approx \sum_{n=0}^{k} \mathbf{Q}^n \frac{(\mu t)^n}{n!} e^{-\mu t} = \sum_{n=0}^{k} \left(\mathbf{I} + \frac{1}{\mu}\mathbf{\Lambda}\right)^n \frac{(\mu t)^n}{n!} e^{-\mu t},$$

where $\mu \geq \max\{\lambda_i\}$. To evaluate this sum, we need to multiply matrices with nonnegative entries, which makes calculations more stable numerically.

Method 2: diagonalization

Suppose $\mathbf{\Lambda}$ is diagonalizable, so there exists a non-singular matrix $\mathbf{T}$ such that $\mathbf{\Lambda} = \mathbf{TDT}^{-1}$, where

$$\mathbf{D} = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_m \end{pmatrix} = \mathrm{diag}(d_1, \ldots, d_m)$$

is a diagonal matrix. Then

$$\mathbf{\Lambda}^k = \mathbf{TDT}^{-1}\mathbf{TDT}^{-1} \cdots \mathbf{TDT}^{-1} = \mathbf{TD}^k\mathbf{T}^{-1},$$

where $\mathbf{D}^k = \mathrm{diag}(d_1^k, \ldots, d_m^k)$. Therefore,

$$\begin{aligned} e^{\mathbf{\Lambda}t} &= \sum_{k=0}^{\infty} \frac{t^k \mathbf{\Lambda}^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{TD}^k\mathbf{T}^{-1}}{k!} = \mathbf{T}\left[\sum_{k=0}^{\infty} \frac{t^k\mathbf{D}^k}{k!}\right]\mathbf{T}^{-1} = \mathbf{T}\begin{pmatrix} \sum_{k=0}^{\infty} \frac{t^k d_1^k}{k!} & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} \frac{t^k d_m^k}{k!} \end{pmatrix}\mathbf{T}^{-1} \\ &= \mathbf{T}\mathrm{diag}(e^{d_1 t}, \ldots, e^{d_m t})\mathbf{T}^{-1} \end{aligned}$$

It is not always possible to analytically determine whether the infinitesimal generator is diagonalizable. Reversible Markov chains are pleasant exceptions. To see this, we need to recall some basic linear algebra.

Definition. The matrices $\mathbf{A}$ and $\mathbf{B}$ are called similar if there exists a nonsingular $\mathbf{P}$ such that $\mathbf{PAP}^{-1} = \mathbf{B}$.

Proposition 2.9. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are similar and* $\mathbf{A}$ *is diagonalizable, then* $\mathbf{B}$ *is diagonalizable.*

Proof. $\mathbf{A}$ is diagonalizable $\Rightarrow$ $\mathbf{A} = \mathbf{TDT}^{-1}$, where $\mathbf{D}$ is a diagonal matrix. From the definition, $\mathbf{B} = \mathbf{PAP}^{-1} = (\mathbf{PT})\mathbf{D}(\mathbf{PT})^{-1}$. □

Also, recall that if $\mathbf{A}$ is symmetric, then $\mathbf{A}$ is diagonalizable.

Proposition 2.10. *Let $\mathbf{\Lambda}$ be the infinitesimal generator of a reversible CTMC with stationary distribution $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$, where $\pi_i > 0$ for all $i \in \Omega$. Then $\mathbf{\Lambda}$ is diagonalizable.*

Proof. Consider a non-singular matrix

$$\mathbf{\Pi} = \begin{pmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_m \end{pmatrix}$$

and $\mathbf{A} = \mathbf{\Pi}^{1/2}\mathbf{\Lambda}\mathbf{\Pi}^{-1/2}$. Let $\mathbf{A} = \{a_{ij}\}$. Then

$$a_{ij} = \pi_i^{1/2}\lambda_{ij}\pi_j^{-1/2} = \pi_i\lambda_{ij}\pi_i^{-1/2}\pi_j^{-1/2} = [\text{detailed balance}] = \pi_j\lambda_{ji}\pi_i^{-1/2}\pi_j^{-1/2} = \pi_j^{1/2}\lambda_{ji}\pi_i^{-1/2} = a_{ji}.$$

Therefore $\mathbf{A}$ is symmetric and diagonalizable as a result. Since $\mathbf{\Lambda}$ is similar to $\mathbf{A}$, $\mathbf{\Lambda}$ is also diagonalizable. □

Example: *Molecular evolution* The Hasegawa-Kimura-Yang (HKY) model describes stochastic mutational process among four nucleotides: A, G, C, T (Hasegawa et al., 1985). The pairs A,G and C,T form two classes, called purines and pyrimidines respectively. Hasegawa and co-authors argued that the biochemistry of nucleotide base interactions suggests that mutations within classes (A $\leftrightarrow$ G, T $\leftrightarrow$ C) should have higher rates than mutations between classes (A,G $\leftrightarrow$ C,T). The two types of mutations are called "transitions" and "transversions." The infinitesimal generator of the HKY CTMC,

$$\mathbf{\Lambda} = \begin{pmatrix} - & \alpha\pi_G & \beta\pi_C & \beta\pi_T \\ \alpha\pi_A & - & \beta\pi_C & \beta\pi_T \\ \beta\pi_A & \beta\pi_G & - & \alpha\pi_T \\ \beta\pi_A & \beta\pi_G & \alpha\pi_c & - \end{pmatrix},$$

is parameterized in terms of a transition rate α, a transversion rate β and a probability vector $\boldsymbol{\pi} = (\pi_A, \pi_G, \pi_C, \pi_T)$. The infinitesimal generator is constructed in such a way that $\boldsymbol{\pi}$ satisfies detailed balance and therefore, $\boldsymbol{\pi}$ is the stationary distribution of this chain. Since the chain is reversible, $\mathbf{\Lambda}$ is diagonalizable. In practice, diagonalization is accomplished by finding the eigenvalues of $\mathbf{\Lambda}$ and the corresponding left and right eigenvectors.

Let $\pi_Y = \pi_C + \pi_T$ and $\pi_R = \pi_A + \pi_G$. Then the eigenvalues of $\mathbf{\Lambda}$ are

$$0, \qquad -\beta, \qquad -(\pi_Y\beta + \pi_R\alpha), \qquad -(\pi_Y\alpha + \pi_R\beta).$$

The corresponding right eigenvectors are

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} -\frac{1}{\pi_R} \\ -\frac{1}{\pi_R} \\ \frac{1}{\pi_Y} \\ \frac{1}{\pi_Y} \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} \frac{\pi_G}{\pi_R} \\ -\frac{\pi_A}{\pi_R} \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_4 = \begin{pmatrix} 0 \\ 0 \\ -\frac{\pi_T}{\pi_Y} \\ \frac{\pi_C}{\pi_Y} \end{pmatrix}.$$

and left eigenvectors are

$$\mathbf{v}_1 = \begin{pmatrix} \pi_A \\ \pi_G \\ \pi_C \\ \pi_T \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -\pi_Y\pi_A \\ -\pi_Y\pi_G \\ \pi_R\pi_C \\ \pi_R\pi_T \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}.$$

Collecting right and left eigenvectors into matrices $\mathbf{U} = (\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \mathbf{u}_4)$ and $\mathbf{V} = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3\ \mathbf{v}_4)$ we get

$$\mathbf{\Lambda} = \mathbf{U}\ \mathrm{diag}[0, -\beta, -(\pi_Y\beta + \pi_R\alpha), -(\pi_Y\alpha + \pi_R\beta)]\mathbf{V}^T$$

and

$$e^{\mathbf{\Lambda}t} = \mathbf{U}\ \mathrm{diag}[1, e^{-\beta t}, e^{-(\pi_Y\beta+\pi_R\alpha)t}, e^{-(\pi_Y\alpha+\pi_R\beta)t}]\mathbf{V}^T.$$

2.2.6 Absorbing continuous-time Markov chains

Consider a continuous-time Markov chain $\{X_t\}$ on a finite state space $\Omega = \{1,\ldots,n\}$. Let

$A = \{1,\ldots,m\}$ be a set of absorbing states and
$B = \{m+1,\ldots,n\}$ be as a set transient states,

meaning that $\lambda_{ij} = 0$ for all $i \in A$ and $j \in B$ and for every $i \in B$ there exists $j \in A$ such that $i \to j$, but $i \not\leftarrow j$. Rearranging the order of the states so that $\Omega = \{A, B\}$, we can write the transition probability matrix in a block matrix form:

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_{AA} & \mathbf{0} \\ \mathbf{\Lambda}_{BA} & \mathbf{\Lambda}_{BB} \end{pmatrix},$$

where $\mathbf{\Lambda}_{AA}$ is an $m \times m$ matrix, $\mathbf{\Lambda}_{BA}$ is an $(n-m)\times m$ matrix, and $\mathbf{\Lambda}_{BB}$ is an $(n-m)\times(n-m)$ matrix.

We are interested in how much time T *that X_t spends in transient states until jumping to one of the closed states*, assuming that $X_0 \in B$. To get a handle of the distribution of such times, we start with the conditional probability of starting in state i and exiting to a set of absorbing states through state j during the interval $[0,t]$,

$$F_{ij}(t) = \Pr(T \le t, X_T = j \,|\, X_0 = i) \text{ for } i \in B \text{ and } j \in A.$$

Since we do not care what happens with the Markov chain after it enters a set of absorbing states, it is convenient to define a new CTMC with infinitesimal generator

$$\bar{\mathbf{\Lambda}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_{BA} & \mathbf{\Lambda}_{BB} \end{pmatrix}.$$

For a new CTMC Y_t governed by the infinitesimal generator $\bar{\mathbf{\Lambda}}$, each closed state becomes an absorbing state. Let $\bar{\mathbf{P}}(t) = e^{\bar{\mathbf{\Lambda}}t} = \{\bar{p}_{ij}(t)\}$. The reason for defining this new process is that for $i \in B$ and $j \in A$,

$$F_{ij}(t) = \Pr(T \le t, X_T = j \,|\, X_0 = i) = \Pr(Y_t = j \,|\, Y_0 = i) = \bar{p}_{ij}(t).$$

Now, let's see if we can block partition $\bar{\mathbf{P}}(t)$ into something sensible. First, notice that

$$\bar{\mathbf{\Lambda}}^n = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{\Lambda}_{BB}^{n-1}\mathbf{\Lambda}_{BA} & \mathbf{\Lambda}_{BB}^n \end{pmatrix} \text{ for } n \ge 1.$$

$$\begin{aligned}
\bar{\mathbf{P}}(t) = e^{\bar{\mathbf{\Lambda}}t} = \sum_{n=0}^{\infty} \frac{\bar{\mathbf{\Lambda}}^n t^n}{n!} &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \sum_{n=1}^{\infty} \mathbf{\Lambda}_{BB}^{n-1}\mathbf{\Lambda}_{BA}\frac{t^n}{n!} & \sum_{n=1}^{\infty}\mathbf{\Lambda}_{BB}^n\frac{t^n}{n!} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \left[\sum_{n=1}^{\infty}\mathbf{\Lambda}_{BB}^{n-1}\frac{t^n}{n!}\right]\mathbf{\Lambda}_{BA} & \sum_{n=0}^{\infty}\mathbf{\Lambda}_{BB}^n\frac{t^n}{n!} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \int_0^t e^{\mathbf{\Lambda}_{BB}u}du\mathbf{\Lambda}_{BA} & e^{\mathbf{\Lambda}_{BB}t} \end{pmatrix}
\end{aligned}$$

Alternatively, we could argue by first noticing the following partition:

$$\bar{\mathbf{P}}(t) = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \bar{\mathbf{P}}_{BA}(t) & \bar{\mathbf{P}}_{BB}(t) \end{pmatrix}$$

and then use the Kolmogorov forward equation, $\bar{\mathbf{P}}'(t) = \bar{\mathbf{P}}(t)\bar{\boldsymbol{\Lambda}}$, that can be rewritten as

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \bar{\mathbf{P}}'_{BA}(t) & \bar{\mathbf{P}}'_{BB}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \bar{\mathbf{P}}_{BA}(t) & \bar{\mathbf{P}}_{BB}(t) \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Lambda}_{BA} & \boldsymbol{\Lambda}_{BB} \end{pmatrix}.$$

Now, we notice that

$$\bar{\mathbf{P}}'_{BB}(t) = \bar{\mathbf{P}}_{BA}(t)\mathbf{0} + \bar{\mathbf{P}}_{BB}(t)\boldsymbol{\Lambda}_{BB} \Rightarrow \bar{\mathbf{P}}_{BB}(t) = e^{\boldsymbol{\Lambda}_{BB}t},$$

because $\bar{\mathbf{P}}_{BB}(0) = \mathbf{I}$. Next, we have

$$\bar{\mathbf{P}}'_{BA}(t) = \bar{\mathbf{P}}_{BA}(t)\mathbf{0} + \bar{\mathbf{P}}_{BB}(t)\boldsymbol{\Lambda}_{BA} = e^{\boldsymbol{\Lambda}_{BB}t}\boldsymbol{\Lambda}_{BA} \Rightarrow \bar{\mathbf{P}}_{BA}(t) = \int_0^t e^{\boldsymbol{\Lambda}_{BB}u} du \boldsymbol{\Lambda}_{BA},$$

because $\bar{\mathbf{P}}_{BA}(0) = \mathbf{0}$.

Now, let's see what we can do with this result. Let $\mathbf{F}(t) = \{F_{ij}(t)\}$ and $\mathbf{f}(t) = \{f_{ij}(t)\}$, where $f_{ij}(t) = F'_{ij}(t)$. Then we have

$$\mathbf{f}(t) = \mathbf{F}'(t) = \bar{\mathbf{P}}'_{oc}(t) = e^{\boldsymbol{\Lambda}_{BB}t}\boldsymbol{\Lambda}_{BA}.$$

Suppose we are interested in the time of absorption of X_t given an initial distribution $\boldsymbol{\mu}^T = (0, \ldots, 0, \mu_{m+1}, \ldots, \mu_n)$. The CDF of this random variable is

$$F_B(t) = \Pr(T \le t) = \sum_{i \in B} \sum_{j \in A} F_{ij}(t)\mu_i = \boldsymbol{\mu}^T_{m+1:n}\mathbf{F}(t)\mathbf{1}.$$

The density of T is

$$f_B(t) = F'_B(t) = \boldsymbol{\mu}^T_{m+1:n}\mathbf{F}'(t)\mathbf{1} = \boldsymbol{\mu}^T_{m+1:n}\mathbf{f}(t)\mathbf{1} = \boldsymbol{\mu}^T_{m+1:n}e^{\boldsymbol{\Lambda}_{BB}t}\boldsymbol{\Lambda}_{BA}\mathbf{1}.$$

So now we can compute the distribution of T if we know $\boldsymbol{\Lambda}$ and $\boldsymbol{\mu}$.

2.2.7 Phase-type distributed random variables

When an absorbing set A consists of just one state, time to absorption is said to follow a phase-type distribution that is characterized by the matrix $\boldsymbol{\Lambda}_{BB}$ and initial distribution $\boldsymbol{\mu}^T = (0, \mu_2, \ldots, \mu_n)$. Important examples of phase-type distribution include exponential distribution, mixture of exponentials, Erlang distribution, hypoexponential distribution, and Coxian distribution. We provide CTMC formulations for these distributions below.

Example: *Phase-type representation of the exponential distribution* When CTMC state spaces consists of two states: one transient and one absorbing, the waiting time until absorption follows an exponential distribution. Therefore, an exponential distribution is a phase-type distribution specified by the infinitesimal rate matrix and initial distribution below.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 \\ \lambda & -\lambda \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Example: *Phase-type representation of a mixture of exponential distributions* Suppose we want to construct a mixture of n exponential distributions with potentially different rates $\lambda_1, \ldots, \lambda_n$ and mixture component probabilities $\mu_1, \ldots, \mu_n$. Draws from this mixture distribution are obtained by first drawing a random index i from a discrete distribution with the probability mass function specified by $\mu_1, \ldots, \mu_n$. Next, we draw a sample from the exponential distribution with rate λ_i. Imagining that the first step step corresponds to randomly choosing an initial transient state of a CTMC, we can represent the mixture of exponentials with a phase-type distribution specified by the infinitesimal rate matrix and initial distribution below. This distribution is sometimes called hyperexponential.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 & & \cdots & 0 \\ \lambda_1 & -\lambda_1 & & & \\ \lambda_2 & & -\lambda_2 & & \\ \vdots & & & \ddots & \\ \lambda_n & & & & -\lambda_n \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

Example: *Phase-type representation of the Erlang distribution* The Erlang distribution is the distribution of a finite sum of iid exponentials. The number of these exponentials k and their common rate λ are the two parameters. We form a CTMC with k transient state and one absorbing state. There are two special transient states: the starting and ending states. The chain always starts in the former. With probability one the chain progresses to another transient state until it reaches the ending state, from which absorption occurs with probability one. The rate of leaving each transient state is λ. The phase-type distribution formulation is given by the infinitesimal rate matrix and initial distribution below.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 & & \cdots & 0 \\ \lambda & -\lambda & & & \\ & \lambda & -\lambda & & \\ & & \ddots & \ddots & \\ & & & \lambda & -\lambda \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Example: *Phase-type representation of the hypoexponential distribution* The hypoexponential distribution extends the Erlang distribution by allowing exponential distributed random variables in the sum to have different rates. In the phase-type formulation, we just need to allow rates of leaving states to be different, as illustrated by the infinitesimal rate matrix and initial distribution below.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 & & \cdots & 0 \\ \lambda_2 & -\lambda_2 & & & \\ & \lambda_3 & -\lambda_3 & & \\ & & \ddots & \ddots & \\ & & & \lambda_n & -\lambda_n \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Example: *Coxian distribution* The Coxian distribution extends the hypoexponential distribution by allowing nonzero rates of absorption from each of the transient states. The the infinitesimal rate matrix and initial distribution below define this phase-type distribution.

$$\boldsymbol{\Lambda} = \begin{pmatrix} 0 & 0 & & \cdots & 0 \\ \lambda_2 & -\lambda_2 & & & \\ \alpha_3 & \lambda_3 & -(\lambda_3+\alpha_3) & & \\ \vdots & & \ddots & \ddots & \\ \alpha_n & & & \lambda_n & -(\lambda_n+\alpha_n) \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Example: *Multistate models in medical statistics* a. Survival models: These are essentially hypo-exponential models

b. Competing risk models:

$$\mathbf{\Lambda} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \lambda_1 & \lambda_2 & -(\lambda_1+\lambda_2) \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

c. Illness-death models: 1 = dead, 2 = healthy, 3 = sick

$$\mathbf{\Lambda} = \begin{pmatrix} 0 & 0 & 0 \\ \lambda_{21} & \lambda_{31} & -(\lambda_{21}+\lambda_{31}) \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31}+\lambda_{32}) \end{pmatrix}, \qquad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ \mu_2 \\ \mu_3 \end{pmatrix}.$$

In medical statistics, it is important to incorporate incorporate covariates into individual transition rates. This can be done, for example, using the following construction:

$$\log\left(\lambda_{ij}^{(k)}\right) = \boldsymbol{\beta}_{ij}^T \mathbf{x}^{(k)},$$

where $\mathbf{x}^{(k)}$ is a vector of covariates corresponding to individual k and $\lambda_{ij}^{(k)}$ are rates of the disease model for this individual.

2.3 Bibliographic remarks

Most of the material in section 2.1 is based on Brémaud (1998). Reversibility, some of the examples, and transient properties of absorbing Markov chains are borrowed from Lange (2004). See Kemeny and Snell (1976) for a detailed treatment of fundamental matrices.

The main sources for general theory in section 2.2 have been Freedman (1983) and Bhattacharya and Waymire (1990).

2.4 Exercises

1. Prove that the Markov property (2.1 is equivalent to each of the following statements:

   (a) Let T_1 be a set of times later than n, and T_0 a set of times less than or equal to n. Let $t_0 = \max T_0$. Then

   $$\Pr(X_k = x_k, k \in T1 | X_l = x_l, l \in T_0) = \Pr(X_k = x_k, k \in T_1 | X_{t_0} = x_{t_0}).$$

   (b) Let T_1 be a set of times later than n, and T_0 a set of times prior to n. Then

   $$\Pr(X_k = x_k, k \in T_1, X_l = x_l, l \in T_0 | X_n = x_n) = \Pr(X_k = x_k, k \in T_1 | X_n = x_n)\Pr(X_l = x_l, l \in T_0 | X_n = x_n).$$

2. (Brémaud 2.1.5) Let $\{X_n\}_{n\geq 0}$ be an homogeneous Markov chain with state space Ω and transition matrix $\mathbf{P}$. Let τ be the first time n for which $X_n \neq X_0$, where $\tau = +\infty$ if $X_n = X_0$ for all $n \geq 0$. Compute $E[\tau | X_0 = i]$ in terms of p_{ii}.