Chapter 2

Computing Distributions

We classify the techniques in two categories. In the first one, the accent is on the exploitation of some form of structure of the problem in order to arrive at exact formulas for the probabilities of interest. This family of methods is called here "exact" even if the complete solution includes sometimes numerical calculation.

The second category of methods is based on numerical linear algbra. The accent there is on the speed (algorithmic complexity) and robustness (numerical stability).

Except when explicitly mentioned, it is assumed that the Markov Chain under consideration has a unique stationary distribution. According to the results of Chapter 1, this happens when the chains are irreducible.

2.1 Exact computations

2.1.1 Stationary distributions

2.1.1.1 Balance equations

According to Theorem 1.4, the stationary distribution of a DTMC, in vector form, is the solution of the linear system:

$$\boldsymbol{\pi}\,\mathbf{P} \;=\; \boldsymbol{\pi}\;,$$

together with the **normalization equation** $\boldsymbol{\pi}.\mathbf{1} = 1$. For a CTMC, the linear system to solve is, according to Theorem 1.14:

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

still with $\boldsymbol{\pi}.\mathbf{1} = 1$.

For each state $i \in \mathcal{E}$, one of the equations of these systems represents the **equilibrium** or **balance** equation for this state. In discrete-time:

$$\sum_{j\in\mathcal{E}} p_{i,j}\pi(i) \;=\; \sum_{j\in\mathcal{E}} \pi(j)p_{j,i}\;. \tag{2.1}$$

In continuous-time:

$$(\sum_{j\neq i} q_{i,j})\pi(i) \;=\; \sum_{j\neq i} \pi(j)q_{j,i}\;.$$

The intepretation of these equations is that, when the Markov Chain is stationary, the "probability flow" into state i is the same as the flow out of state i. These balance equations are interpreted as flow conservation equations.

This conservation law can be generalized to any subset of $\mathcal{E}$, which gives a property very useful for computations.

Theorem 2.1 (Global Balance Equations)**.** *Assume a Markov Chain over $\mathcal{E}$ has a stationary distribution π. For any $S \subset \mathcal{E}$, if $\overline{S}$ is the complement of S in $\mathcal{E}$, there holds:*

- *if the chain is in discrete time:*

$$\sum_{i \in S, j \in \overline{S}} \pi(i)\ P_{i,j} \ = \ \sum_{i \in \overline{S}, j \in S} \pi(i)\ P_{i,j}\ .$$

- *if the chain is in continuous time:*

$$\sum_{i \in S, j \in \overline{S}} \pi(i)\ q_{i,j} \ = \ \sum_{i \in \overline{S}, j \in S} \pi(i)\ q_{i,j}\ .$$

Proof. We prove the result for discrete-time chains. The proof for continuous-time ones is similar. Starting from the individual state balance equations, one has the series of transformations:

$$\begin{aligned}
\sum_{i \in \mathcal{S}, j \in \overline{\mathcal{S}}} \pi(i) P_{i,j} &= \sum_{i \in \mathcal{S}} \pi(i) \sum_{j \in \overline{\mathcal{S}}} P_{i,j} && (2.2)\\
&= \sum_{i \in \mathcal{S}} \pi(i) \left(1 - \sum_{j \in \mathcal{S}} P_{i,j}\right) && (2.3)\\
&= \sum_{i \in \mathcal{S}} \pi(i) - \sum_{i \in \mathcal{S}, j \in \mathcal{S}} P_{i,j} && (2.4)\\
&= \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{E}} \pi(j) P_{j,i} - \sum_{i \in \mathcal{S}, j \in \mathcal{S}} P_{i,j} && (2.5)\\
&= \sum_{i \in \mathcal{S}} \left(\sum_{j \in \mathcal{S}} \pi(j) P_{j,i} + \sum_{j \in \overline{\mathcal{S}}} \pi(j) P_{j,i} \right) - \sum_{i \in \mathcal{S}, j \in \mathcal{S}} P_{i,j} \\
&= \sum_{i \in \mathcal{S}, j \in \mathcal{S}} \pi(j) P_{j,i} + \sum_{i \in \mathcal{S}, j \in \overline{\mathcal{S}}} \pi(j) P_{j,i} - \sum_{i \in \mathcal{S}, j \in \mathcal{S}} P_{i,j} \\
&= \sum_{j \in \overline{\mathcal{S}}, i \in \mathcal{S}} \pi(j) P_{j,i}\ . && (2.6)
\end{aligned}$$

The passage from (2.2) to (2.3) uses the fact that transition probabilities from state i sum up to 1. The passage from (2.4) to (2.5) uses (2.1). The result is equivalent to (2.6) with a change of summation indices. □

2.1.1.2 Application: birth and death processes

We first illustrate the use of Theorem 2.1 to the case of **birth and death** processes.

A birth and death process is a CTMC on the state space $\mathcal{E} = \mathbb{N}$ (or an interval of it) where transitions occur only between neighboring states n and $n \pm 1$. The transition upwards is interpreted as a birth, the transition downwards as a death, hence the name of the process.

The evolution of the state (population) $X(t)$ is according to the transition rates:

$$\begin{array}{lcll} n & \rightarrow & n+1 & \text{with rate } \lambda_n \\ n & \rightarrow & n-1 & \text{with rate } \mu_n \quad n > 0 \end{array}$$

In other words: the entries of the infinitesimal generator $\mathbf{Q}$ are:

$$Q_{n,n+1} = \lambda_n \ , \qquad Q_{n,n-1} = \mu_n \ , \qquad Q_{n,n} = -(\lambda_n + \mu_n)$$

and all other entries are 0. The transition diagram of this process is illustrated in Figure 2.1.1.2.

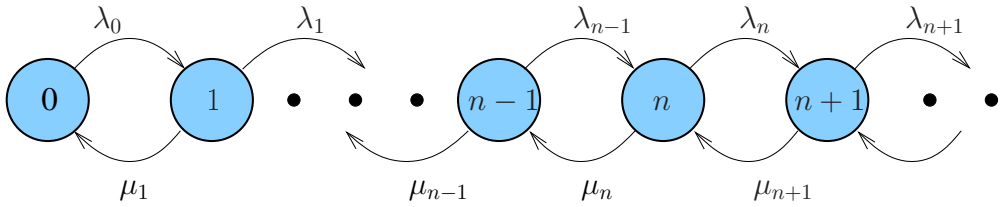

Figure 2.1: The general birth-death process

Let us compute the stationary probabilities. The equilibrium equations for the stationary probabilities $\pi(n)$ are:

$$\underbrace{(\lambda_n + \mu_n)\ \pi(n)}_{\text{output flow}} \ = \ \underbrace{\lambda_{n-1}\pi(n-1) \ + \ \mu_{n+1}\pi(n+1)}_{\text{input flow}} \ .$$

provided that $n \geq 1$. This set of equations has to be solved for the unknowns $\pi(n), n \in \mathbb{N}$.

This task is simplified by using the global balance equations. Take the set $S = \{0, 1, \ldots, n\}$. Applying Theorem 2.1 to this set, and using the fact that only one transition goes out of S and only one goes into S, we obtain:

$$\lambda_n \pi(n) \ = \ \mu_{n+1}\pi(n+1) \ .$$

This is a simple recurrence, the solution of which is: (to be corrected)

$$\pi(n) \ = \ \pi(0) \ \prod_{i=1}^{n} \ \frac{\lambda_{i-1}}{\mu_i} \ . \tag{2.7}$$

There remains one unknown value: $\pi(0)$. It is obtained using the normalization equation: $1 = \sum_n \pi(n)$. This leads to:

$$\pi(0) \ \left[\sum_{n=0}^{\infty} \ \prod_{i=1}^{n} \ \frac{\lambda_{n-1}}{\mu_n} \right] \ = \ 1 \ . \tag{2.8}$$

Then, two situations occur. Either the term inside brackets is finite. Let G be its value. Then $\pi(0) = 1/G$. In this case, there is a summable solution to the system $\pi\mathbf{Q} = 0$, and according to Theorem 1.10, the chain is ergodic.

Or else, the series in the term in brackets diverges to infinity. In that case there is no stationary distribution, still according to Theorem 1.10.

2.1.1.3 The constant-rate birth-death

The most used birth and death process is the one with constant birth rate and constant death rate. It corresponds to the model of the previous section with the particular values:

$$\lambda_n = \lambda \qquad \mu_n = \mu$$

for all n. The transition diagram is illustrated in Figure 2.2.

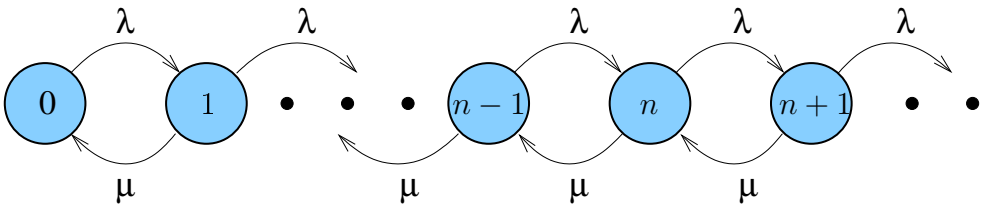

Figure 2.2: Constant-rate birth-death process

The expression (2.7) specialized to this case gives:

$$\pi(n) \;=\; \pi(0)\;\left(\frac{\lambda}{\mu}\right)^n \;. \tag{2.9}$$

The unknown value $\pi(0)$ is given by (2.8) which reduces here to:

$$1 \;=\; \pi(0)\;\left[\sum_{n=0}^{\infty}\;\left(\frac{\lambda}{\mu}\right)^n\right] \;=\; \frac{\pi(0)}{1-\lambda/\mu}\;. \tag{2.10}$$

The expression above is valid only if $\lambda/\mu < 1$. In that case, it provides indeed the value $\pi(0) = 1 - \lambda/\mu$. Gathering the elements of solution, we finally find that the stationary distribution exists and is given by, for all $n \in \mathbb{N}$,

$$\pi(n) \;=\; \left(1-\frac{\lambda}{\mu}\right)\;\left(\frac{\lambda}{\mu}\right)^n \;.$$

This is a **geometric distribution**. The result holds under the condition that $\lambda < \mu$, in other words, that the birth date be smaller than the death rate. Such a condition is called an **ergodicity condition**.

If this condition does not hold, there is no stationary distribution. The precise analysis of the case (out of the scope of this document) leads to the conclusion that: if $\lambda = \mu$, the chain is *null recurrent*. If $\lambda > mu$, it is *transient*.

2.1.1.4 Discrete-time birth and death processes

The discrete-time version of the process studied in Section 2.1.1.3 is represented in Figure 2.3. Such a process is sometimes called a "random walk", or the "drunkard's walk". The image is that of a particle (or a drunk man) who wanders in the street, choosing the direction "at random" at each step.

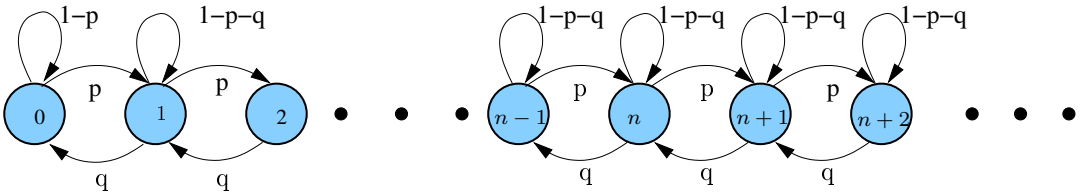

Figure 2.3: A random walk on $\mathbb{N}$

Here, p and q are the probabilties of a birth and a death, respectively. It is required that $p+q \leq 1$. The remaining probability $1-p-q$ is the probability of *status quo*. Observe the specific value for the probability of staying in state 0.

The Global Balance Equations applied to the set $\mathcal{S} = \{0, \ldots, n\}$ gives, for all $n \in \mathbb{N}$:

$$p\ \pi(n) \ = \ q\ \pi(n+1)\ ,$$

and as a consequence, for all n:

$$\pi(n) \ = \ \pi(0)\ \left(\frac{p}{q}\right)^n \ .$$

As in Section 2.1.1.3, two cases have to be considered. Either $p < q$ and there exists a stationary distribution given by:

$$\pi(n) \ = \ \left(1-\frac{p}{q}\right)\ \left(\frac{p}{q}\right)^n \ .$$

This is again a **geometric distribution**. Else $p \geq q$, and there is no stationary distribution.

2.1.1.5 Tree-like chains

When the transition diagram has a structure such that the one displayed in Figure 2.4, Global Balance Equations can also be used to provide a simple solution.

Indeed: consider two states s and t which are connected in the diagram. If one removes the two transitions $s \to t$ and $t \to s$, the tree is disconnected in to exactly two connected components. Let $\mathcal{S}$ be one of them. Then the Global Balance Equation applied to this set gives:

$$\pi(s)p_{s,t} \ = \ \pi(t)p_{t,s}\ .$$

Next, pick a particular state Ω (the "root" of the tree). For any state s, there is a unique path going from Ω to s. Let us denote it with $\gamma(s) = (\Omega = s_0, \ldots, s_m = s)$. The length of this path $|\gamma(s)|$ (still

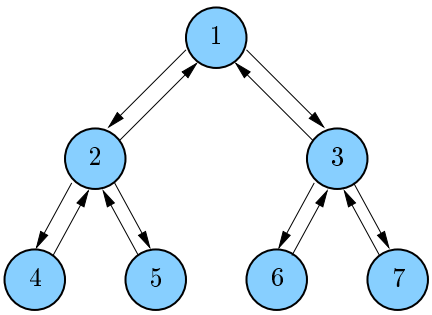

Figure 2.4: A tree-like chain

measured in number of jumps) is called the *depth* of s with respect to the root Ω. Applying the equation above, it is clear that:

$$\pi(s) \;=\; \pi(\Omega) \prod_{i=1}^{|\gamma(s)|} \frac{p_{s_{i-1},s_i}}{p_{s_i,s_{i-1}}} \;.$$

The unknown value $\pi(\Omega)$ is obtained by normalization:

$$\pi(\Omega) \;=\; \left[\sum_{s\in\mathcal{E}} \prod_{i=1}^{|\gamma(s)|} \frac{p_{s_{i-1},s_i}}{p_{s_i,s_{i-1}}} \right]^{-1}$$

provided the series converges.

2.1.1.6 Stationary probabilities as recurrences

When the transitions have a regular structure, the solution can sometimes be computed using recurrences. These recurrences in turn can have an explicit solution. We illustrate the principle through a CTMC example. This principle applies to more complicated examples, and also to DTMC.

Consider for instance the chain depicted in Figure 2.5.

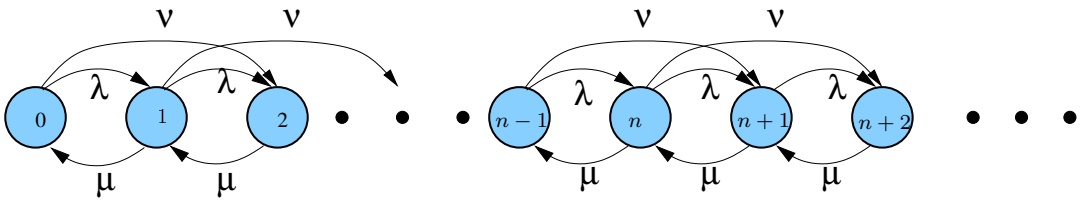

Figure 2.5: A CTMC with regular structure

The state balance equations are the following:

$$(\lambda+\nu)\pi(0) \;=\; \mu\pi(1) \tag{2.11}$$

$$(\lambda+\nu+\mu)\pi(1) \;=\; \mu\pi(2) \tag{2.12}$$

$$(\lambda+\nu+\mu)\pi(n) \;=\; \mu\pi(n+1) \;+\; \nu\pi(n-2)\;, \qquad n\geq 2\;. \tag{2.13}$$

This last equation can be seen as a recurrence allowing to compute $\pi(n+1)$ as a function of $\pi(n)$ and $\pi(n-2)$. This is a linear recurrence with constant coefficient, of order 3.

If one applies the Global Balance Equations to the set $\mathcal{S} = \{0, \ldots, n\}$, one gets instead: The state balance equations are the following: Equation (2.11) and

$$(\lambda + \nu)\pi(n) + \nu\pi(n-1) \;=\; \mu\pi(n+1)\;, \qquad n \geq 1\;. \tag{2.14}$$

This is still a linear recurrence with constant coefficients, but it is of order 2.

It is known that the general solution of Recurrence (2.14) is of the form:

$$\pi(n) \;=\; A\alpha_1^n \;+\; B\alpha_2^n\;,$$

where α_1 and α_2 are the two roots of the *characteristic equation*

$$\mu\alpha^2 - (\lambda + \nu)\alpha - \nu \;=\; 0\;.$$

The solution of the problem of computing $\boldsymbol{\pi}$ is complete if we can determine the constants A and B. To that end, we use two equations not used yet: the balance equation (2.11) and the normalization equation $\boldsymbol{\pi}\mathbf{1} = 1$. Given the general form for $\pi(n)$ and *if* $|\alpha_1| < 1$ and $|\alpha_2| < 1$: we must have:

$$1 \;=\; \frac{A}{1-\alpha_1} \;+\; \frac{B}{1-\alpha_2}\;.$$

Equation (2.11) writes as:

$$(\lambda + \nu)(A + B) \;=\; \mu(A\alpha_1 + B\alpha_2)\;.$$

The analysis of the characteristic polynomial yields the following facts: one of its roots α_1 is always in the interval $(-1, 0)$ if $\nu > 0$, and is 0 if $\nu = 0$. The other root α_2 is positive, and $\alpha_2 < 1$ if and only if $\mu > \lambda + 2\nu$. If this stability condition is not satisfied, there is no stationary distribution.

Assuming therefore $\mu > \lambda + 2\nu$, the final expression for the solution can be expressed as follows:

$$\begin{aligned}
\pi(n) &= A\alpha_1^n \;+\; B\alpha_2^n \\
A &= \frac{\mu - 2\nu - \lambda}{2\mu}\,(1 - \frac{\lambda + \nu}{\Delta}) \\
B &= -\,A\,\frac{\lambda + \nu - \mu\alpha_1}{\lambda + \nu - \mu\alpha_2} \\
\Delta &= \sqrt{(\lambda + \nu)^2 + 4\mu\nu} \\
\alpha_1 &= \frac{\lambda + \nu - \Delta}{2\mu} \\
\alpha_2 &= \frac{\lambda + \nu + \Delta}{2\mu}\;.
\end{aligned}$$

This solution reduces to that of Section 2.1.1.3 if $\nu = 0$.

2.1.1.7 Reversibility and truncation

If $\{X(t); t \in \mathbb{R}\}$ is a stationary CTMC, its **reversed chain** is the process $\{X(-t); t \in \mathbb{R}\}$. It is also a stationary CTMC.

Let $\mathbf{Q}$ be the infinitesimal generator of the chain $\{X(t)\}$, and let $\boldsymbol{\pi}$ be a stationary distribution. Then the infinitesimal generator of the reversed chain, say $\mathbf{Q}'$, is given by:

$$Q'_{i,j} \;=\; Q_{j,i}\,\frac{\pi(j)}{\pi(i)}\;.$$

It can be verified that $\boldsymbol{\pi}$ is also a stationary distribution for the reversed chain. However, the generators do not coincide in general. They do for a special class of Markov chains that we now introduce.

Definition 2.1 (Reversible Markov Chains)**.** *A continuous-time Markov chain is reversible if* $\{X(t)\}$ *have* $\{X(-t)\}$ *the same generator, that is, if for all pair of states* $(i,j) \in \mathcal{E} \times \mathcal{E}$*,*

$$\pi(i)\;q_{i,j} \;=\; \pi(j)\;q_{j,i}\;. \tag{2.15}$$

A CTMC is said to satisfy the **local balance** condition if for any transition $j \leftrightarrow k$:

$$\pi(j)\;q_{j,k} \;=\; \pi(k)\;q_{k,j}\;.$$

This implies that when transition $j \to k$ exists, transition $k \to j$ must exist as well.

The class of Reversible Markov chains has property very useful for computing distributions: it is stable with respect to *truncation*.

Definition 2.2 (Truncated Markov Chain)**.** *Consider a CTMC over a countable state space* $\mathcal{E}$ *and generator* $\mathbf{Q}$*. Let* $\mathcal{A} \subset \mathcal{E}$*. The truncated Markov chain over the sub-space* $\mathcal{A}$ *is the CTMC with generator* $\mathbf{Q}_{\mathcal{A}}$ *defined as:*

$$(q_{\mathcal{A}})_{jk} \;=\; \begin{cases} q_{jk} & \text{if } j,k \in \mathcal{A} \\ 0 & \text{otherwise.} \end{cases} \tag{2.16}$$

In words, the transitions of the process are the same as the original process as long as the state is in $\mathcal{A}$. But transitions outside $\mathcal{A}$ have rate 0. If the process starts inside $\mathcal{A}$, it stays there forever, and it can be considered as a CTMC on the space $\mathcal{A}$.

Theorem 2.2 (Stationary distribution of truncated reversible chains)**.** *Let* $\mathcal{A} \subset \mathcal{E}$*. Consider some reversible CTMC, with stationary distribution* $\boldsymbol{\pi}$*. Then the truncated Markov chain over* $\mathcal{A}$ *is also reversible, and its stationary distribution is given by: for* $i \in \mathcal{A}$

$$\pi(i)/\big(\sum_{j\in\mathcal{A}} \pi(j)\big)\;.$$

We illustrate this principle on some examples. Consider first the Birth and Death process of Section 2.1.1.2. This CTMC is reversible because, as we have seen, it satisfies the local balance equation:

$$\lambda_n \pi(n) \;=\; \mu_{n+1}\pi(n+1)\;.$$

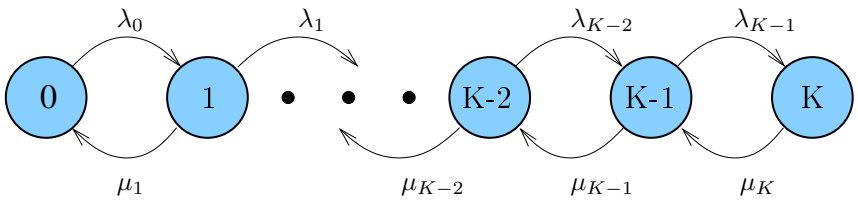

Figure 2.6: The truncated Birth and Death Process

Consider now the same process, truncated to the set $\mathcal{E}' = \{0, \ldots, K\}$. The diagram of the resulting CTMC is represented in Figure 2.6.

Applying Theorem 2.2, we obtain directly the result that the stationary distribution of this truncated chain is of the form

$$\pi(n) \;=\; G \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} \;,$$

for some normalization constant G. This result is however obvious using *global* balance equations as in Section 2.1.1.2.

As a second, less obvious example, consider the CTMC displayed in Figure 2.7 (left), which evolves in the state space $\mathcal{E} = \mathbb{N}^2$. This chain is also reversible, with stationary distribution:

$$\pi(i,j) \;=\; \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\alpha}{\beta}\right) \left(\frac{\alpha}{\beta}\right)^j \;.$$

Indeed, it can be checked using this formula that the transition rates satisfy the two families of local balance equations: for all $(i,j) \in \mathbb{N} \times \mathbb{N}$,

$$\begin{aligned} \pi(i,j)\ \lambda &= \pi(i+1,j)\ \mu \\ \pi(i,j)\ \alpha &= \pi(i,j+1)\ \beta \;. \end{aligned}$$

Next, consider the chain with the same transitions but truncated to the state space $\mathcal{E}' = \{(i,j) \in \mathbb{N} \times \mathbb{N} \mid i \leq 3, j \leq 3, i+j \leq 4\}$, by simply removing transitions that go outside this set. The diagram of this chain is in Figure 2.7 (right). Applying Theorem 2.2, we find directly that the stationary distribution of this truncated chain is of the form: for every $(i,j) \in \mathcal{E}'$,

$$\pi(i,j) \;=\; G \left(\frac{\lambda}{\mu}\right)^i \left(\frac{\alpha}{\beta}\right)^j \;,$$

where G is a normalization constant such that $\sum_{(i,j)\in\mathcal{E}'} \pi(i,j) = 1$.

2.1.2 Transient probabilities

According to Theorem 1.3, the transition probabilities and state probabilities of a DTMC can be computed using the powers of the transition matrix $\mathbf{P}$. Section A.3.2, in particular Theorem A.2,

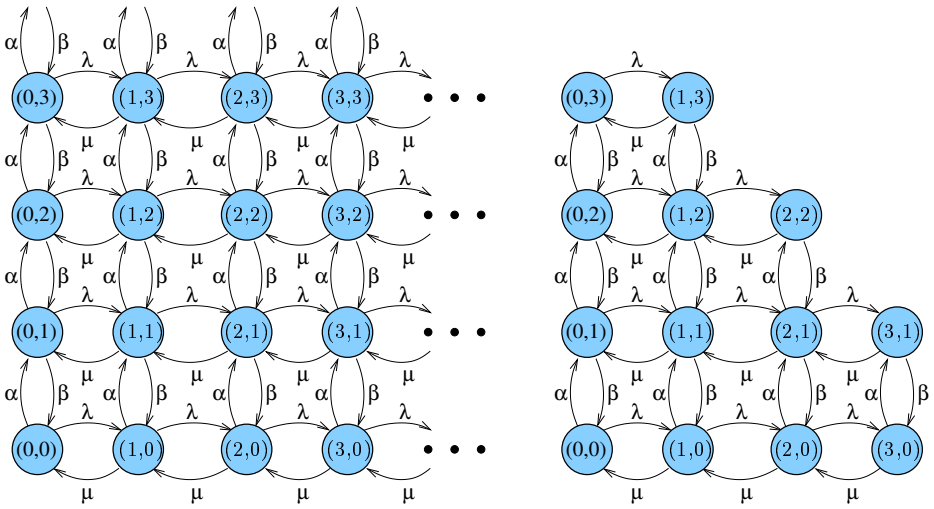

Figure 2.7: A two-dimensional reversible CTMC (left) and a truncated version (right)

provides general formulas for powers of matrices. In particular, if $\mathbf{P}$ is diagonalizable, there exist rank-one matrices A_i such that:

$$\mathbf{P}^m \;=\; \sum_{i=1}^{n} \lambda_i^m \; A_i \;,$$

where the λ_i are the eigenvalues of matrix $\mathbf{P}$.

Similarly, according to Theorem 1.13, the transition probabilities and the transient state probabilities can be computed using the exponential of the infinitesimal generator: $e^{t\mathbf{Q}}$. We look now for ways to obtain explicit expression for this matrix.

If $\mathbf{Q}$ is diagonalizable, there exist matrices $\mathbf{R}$, $\mathbf{S} = \mathbf{R}^{-1}$ and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ such that:

$$\mathbf{Q} \;=\; \mathbf{S}\,\mathbf{D}\,\mathbf{R}\;.$$

Then:

$$e^{t\mathbf{Q}} \;=\; \mathbf{S} \begin{pmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_n t} \end{pmatrix} \mathbf{S}^{-1}\;. \tag{2.17}$$

As a consequence, we have the *spectral decomposition*: there exist matrices (with rank 1) $A_1, \ldots, A_n$ such that:

$$e^{t\mathbf{Q}} \;=\; \sum_{i=1}^{n} e^{\lambda_i t} \; A_i \;.$$

It is therefore possible to obtain explicit formulas for $\mathbf{P}^m$ or $e^{t\mathbf{Q}}$, and $\pi(m)$ or $\pi(t)$, if it is possible to explicitly diagonalize the matrix $\mathbf{P}$ or $\mathbf{Q}$.

In the following paragraph, we study several such cases.

2.1.2.1 The two-state DTMC

Let $\{X_n, n = 0, 1, \ldots\}$ be a DTMC with state space $\mathcal{E} = \{0, 1\}$ and transition matrix:

$$\mathbf{P} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix} \quad 0 \leq \alpha, \beta \leq 1 \ .$$

Define, for all $n = 0, 1, ...$, the vector $\pi_n = (\pi_{n,0}, \pi_{n,1}) = (P(X_n = 0), P(X_n = 1))$.

The eigenvalues of $\mathbf{P}$ are 1 and $\alpha + \beta - 1$.

First, get rid of the case where the two eigenvalues are equal: $\alpha + \beta = 2$ that is, $\alpha = \beta = 1$. Then, $P = I$ and P is already diagonal.

When $\alpha + \beta \neq 2$, define

$$\mathbf{S} = \begin{pmatrix} 1 & 1-\alpha \\ 1 & \beta - 1 \end{pmatrix} \ ,$$

one has:

$$\mathbf{P} \ = \ \mathbf{SDS}^{-1} \ = \ \begin{pmatrix} 1 & 1-\alpha \\ 1 & \beta-1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \alpha+\beta-1 \end{pmatrix} \begin{pmatrix} \beta-1 & \alpha-1 \\ -1 & 1 \end{pmatrix} \frac{1}{\alpha+\beta-2} \ .$$

As a consequence, the spectral decomposition gives:

$$\mathbf{P}^n \ = \ \frac{1}{\alpha+\beta-2} \begin{pmatrix} \beta-1 & \alpha-1 \\ \beta-1 & \alpha-1 \end{pmatrix} \ + \ \frac{(\alpha+\beta-1)^n}{\alpha+\beta-2} \begin{pmatrix} \alpha-1 & 1-\alpha \\ 1-\beta & \beta-1 \end{pmatrix} \ .$$

Note the case $\alpha + \beta = 1$ where $\mathbf{P}^n = \mathbf{P}$ for all n. Also, if $\alpha + \beta = 0$, one can get directly: $\mathbf{P}^n = \mathbf{P}$ for n odd and $\mathbf{P}^n = I$ for n even. If $\alpha + \beta = 2$, the formula above does not hold, but $\mathbf{P}^n = I$ for all n.

The state probabilities are obtained as follows. When $\alpha + \beta = 2$, $\pi_n = \pi_0$ for all n, of course.

In the case $\alpha + \beta < 2$, the above expression and the relation $\pi_n = \pi_0 \mathbf{P}^n$ give, for all $n \in \mathbb{N}$:

$$\begin{aligned} \pi_{n,0} &= \frac{1}{\alpha+\beta-2} \left[\beta - 1 + (\alpha+\beta-1)^n((\alpha-1)\pi_{0,0} + (1-\beta)\pi_{0,1})\right] \\ \pi_{n,1} &= \frac{1}{\alpha+\beta-2} \left[\alpha - 1 - (\alpha+\beta-1)^n((\alpha-1)\pi_{0,0} + (1-\beta)\pi_{0,1})\right] \ . \end{aligned}$$

The sequences $\pi_{n,0}$ and $\pi_{n,1}$ converge if and only if $-1 < \alpha + \beta - 1 \leq 1$, i.e. if and only if $\alpha > 0$ or $\beta > 0$. The limiting steady state vector is then:

$$\lim_{n\to\infty} \boldsymbol{\pi}_n \ = \ \boldsymbol{\pi} \ = \ \left(\frac{1-\beta}{2-\alpha-\beta}, \frac{1-\alpha}{2-\alpha-\beta}\right) \ .$$

In the last case $\alpha + \beta = 0$, the chain is irreducible, but periodic and has no steady state either.

2.1.2.2 The two-state CTMC

Let $\{X(t), t \geq 0\}$ be a CTMC with state space $\mathcal{E} = \{0, 1\}$ and infinitesimal generator:

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \qquad \lambda, \mu > 0 \ .$$

Define, for all $n = 0, 1, ...$, the vector $p(t) = (p_0(t), p_1(t)) = (P(X(t) = 0), P(X(t) = 1))$.

The eigenvalues of Q are 0 with right eigevector $(1, 1)^t$ and $-(\lambda + \mu)$ with eigenvector $(-\lambda, \mu)^t$. The eigenvalues are always distinct because $\lambda > 0$ and $\mu > 0$. One gets:

$$\mathbf{Q} \;=\; \begin{pmatrix} 1 & -\lambda \\ 1 & \mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -(\lambda+\mu) \end{pmatrix} \begin{pmatrix} \mu & \lambda \\ -1 & 1 \end{pmatrix} \frac{1}{\lambda+\mu} \; .$$

The spectral expansion is:

$$e^{\mathbf{Q}t} \;=\; \frac{1}{\lambda+\mu} \begin{pmatrix} \mu & \lambda \\ \mu & \lambda \end{pmatrix} \;+\; \frac{e^{-(\lambda+\mu)t}}{\lambda+\mu} \begin{pmatrix} \lambda & -\lambda \\ -\mu & \mu \end{pmatrix} \; .$$

From the above expression, and $p(t) = p(0)e^{\mathbf{Q}t}$, one has:

$$\begin{aligned} p_0(t) &= \frac{1}{\lambda+\mu} \Big(\mu + e^{-(\lambda+\mu)t} (\lambda p_0(0) - \mu p_1(0) \Big) \\ p_1(t) &= \frac{1}{\lambda+\mu} \Big(\lambda - e^{-(\lambda+\mu)t} (\lambda p_0(0) - \mu p_1(0) \Big) \; . \end{aligned}$$

As $\lambda + \mu > 0$, the functions $p_0(t)$ and $p_1(t)$ always have a limit when t goes to infinity;

$$\lim_{t\to\infty} p(t) \;=\; p \;=\; \left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu} \right) \; .$$

This is the stationary distribution of the CTMC. Note that the convergence is exponentially fast.

2.1.2.3 Felsenstein-81 generators

The model of Felsenstein is used in Genetics. Felsenstein-type matrices (F81 for short, see [6]) are such that

$$q_{i,j} \;=\; \mu p_j \;, i \neq j \;, \qquad q_{i,i} \;=\; -\mu(1 - p_i) \;,$$

where $\boldsymbol{\pi} = (p_1, \ldots, p_N)$ is a probability distribution, and $\mu > 0$ a rate parameter. Algebraically, we have:

$$\mathbf{Q} \;=\; \mu \left(\mathbf{1}\boldsymbol{\pi} \;-\; \mathbf{I} \right) \; .$$

These matrices are constant-speed matrices.

The stationary distribution of this generator is, by construction, $\boldsymbol{\pi}$.

The analysis of the spectrum of $\mathbf{Q}$ yields the following facts:

- the eigenvalues are $\alpha_1 = 0$, and $\alpha_2 = \ldots = \alpha_N = -\mu$;
- left eigenvectors are $S_1 = \boldsymbol{\pi}$ and for $2 \leq i \leq N$: $S_i = (0, \ldots, 0, 1, -1, 0, \ldots, 0)$, where the "1" is in position $i - 1$;
- the system of right eigenvectors compatible with the above left eigenvectors is $R_1 = \mathbf{1}$ and for $2 \leq i \leq N$:

$$R'_i \;=\; (\underbrace{1, 1, \ldots, 1}_{i-1 \text{ times}}, 0, \ldots, 0) \;-\; \left(\sum_{j=1}^{i-1} p_j \right) \; \mathbf{1} \; .$$

For instance with $N = 4$, we have:

$$\mathbf{Q} = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\mu & 0 & 0 \\ 0 & 0 & -\mu & 0 \\ 0 & 0 & 0 & -\mu \end{pmatrix} \begin{pmatrix} 1 & p_2+p_3+p_4 & p_3+p_4 & p_4 \\ 1 & -p_1 & p_3+p_4 & p_4 \\ 1 & -p_1 & -p_1-p_2 & p_4 \\ 1 & -p_1 & -p_1-p_2 & -p_1-p_2-p_3 \end{pmatrix} .$$

Using this decomposition, we obtain the formula:

$$e^{\mathbf{Q}t} \;=\; e^{-\mu t}\;\mathbf{I} \;+\; (1-e^{-\mu t})\;\mathbf{1}\boldsymbol{\pi}\;. \tag{2.18}$$

However, in this specific case, there is a direct way to obtain this result. First, since the matrix $\mathbf{I}$ commutes with any matrix, we have (see Appendix A.4):

$$e^{\mathbf{Q}t} \;=\; e^{\mu t(\mathbf{1}\boldsymbol{\pi}\,-\,\mathbf{I})} \;=\; e^{\mu t\mathbf{1}\boldsymbol{\pi}\,-\,\mu t\mathbf{I}} \;=\; e^{\mu t\mathbf{1}\boldsymbol{\pi}}\;e^{-\mu t}\;.$$

Next, from the definition of the exponential:

$$e^{\mu t\mathbf{1}\boldsymbol{\pi}} \;=\; \sum_{m=0}^{\infty}\frac{(\mu t)^m}{m!}(\mathbf{1}\boldsymbol{\pi})^m\;.$$

But $(\mathbf{1}\boldsymbol{\pi})^0 = \mathbf{I}$ and $(\mathbf{1}\boldsymbol{\pi})^m = \mathbf{1}\boldsymbol{\pi}$ for all $m > 0$. Consequently,

$$e^{\mu t\mathbf{1}\boldsymbol{\pi}} \;=\; \mathbf{I} \;+\; \mathbf{1}\boldsymbol{\pi}\;\sum_{m=1}^{\infty}\frac{(\mu t)^m}{m!} \;=\; \mathbf{I} \;+\; \mathbf{1}\boldsymbol{\pi}\;(e^{\mu t}-1)\;.$$

This gives (2.18).

2.1.2.4 Felsenstein-Churchill-96 transition matrices

A discrete-time analogue of the model of Section 2.1.2.3 has been proposed in [7] in Genetics. Assume that the transition probabilities are such that:

$$p_{i,j} \;=\; (1-\lambda)p_i\;, i \neq j\;, \qquad p_{i,i} \;=\; \lambda + (1-\lambda)p_i\;,$$

where $\boldsymbol{\pi} = (p_1, \ldots, p_N)$ is a probability distribution, and $\lambda \in [0,1]$. Algebraically, we have:

$$\mathbf{P} \;=\; \lambda\mathbf{I} + (1-\lambda)\mathbf{1}\boldsymbol{\pi}\;.$$

The stationary distribution of this generator is, by construction, $\boldsymbol{\pi}$. By recurrence, the powers of $\mathbf{P}$ are computed according to:

$$\mathbf{P}^m \;=\; \lambda^m\mathbf{I} + (1-\lambda^m)\mathbf{1}\boldsymbol{\pi}\;.$$

2.1.2.5 Tamura-Nei-93 generators

Tamura-Nei matrices (TN93 for short, see [25]) are 4×4 matrices with the following structure:

$$\mathbf{Q} \;=\; \begin{pmatrix} - & \pi_C & \kappa_R\pi_G & \pi_T \\ \pi_A & - & \pi_G & \kappa_Y\pi_T \\ \kappa_R\pi_A & \pi_C & - & \pi_T \\ \pi_A & \kappa_Y\pi_C & \pi_G & - \end{pmatrix} ,$$

where $\boldsymbol{\pi} = (\pi_A, \pi_C, \pi_G, \pi_T)$ is a probability distribution, $\kappa_R > 0$ and $\kappa_Y > 0$ rate parameters. The diagonal terms are such that $\mathbf{Q}$ is an infinitesimal generator. The stationary distribution of this generator is $\boldsymbol{\pi}$. Special cases of this family include Hasegawa-Kishino-Yano-85 matrices [8] when $\kappa_R = \kappa_Y = \kappa$, and Felsenstein-81 [6] matrices when $\kappa_R = \kappa_Y = 1$.

The diagonalization of $\mathbf{Q}$ yields the following decomposition:

$$\mathbf{Q} \;=\; \mathbf{R}\,\mathbf{D}\,\mathbf{S} \;=\; \begin{pmatrix} 0 & -\pi_G & -\pi_Y & 1 \\ -\pi_T & 0 & \pi_R & 1 \\ 0 & \pi_A & -\pi_Y & 1 \\ \pi_C & 0 & \pi_R & 1 \end{pmatrix} \mathbf{D} \begin{pmatrix} 0 & -\frac{1}{\pi_Y} & 0 & \frac{1}{\pi_Y} \\ -\frac{1}{\pi_R} & 0 & \frac{1}{\pi_R} & 0 \\ -\frac{\pi_A}{\pi_R} & \frac{\pi_C}{\pi_Y} & -\frac{\pi_G}{\pi_R} & \frac{\pi_T}{\pi_Y} \\ \pi_A & \pi_C & \pi_G & \pi_T \end{pmatrix} , \tag{2.19}$$

where $\mathbf{D} = \mathrm{diag}(-(\pi_R + \kappa_Y\pi_Y), -(\kappa_R\pi_R + \pi_Y), -1, 0)$, and with the notation $\pi_R = \pi_A + \pi_G$ and $\pi_Y = \pi_C + \pi_T$.

Next,

$$e^{t\mathbf{Q}} \;=\; \mathbf{1}\boldsymbol{\pi} \;+\; e^{-t(\pi_R+\kappa_Y\pi_Y)}\, R_1S_1 \;+\; e^{-t(\kappa_R\pi_R+\pi_Y)}\, R_2S_2 \;+\; e^{-t}R_3S_3 \;.$$

The vectors R_i are the columns of the matrix $\mathbf{R}$ given in (2.19). The vectors S_i are the rows of $\mathbf{S}$ in the same formula. A detailed closed-form expression for transition probabilities is given below. The convention used in these formulas is that states belong to pairs $R = \{A, C\}$ and $Y = \{G, T\}$, and that if i is a state, i' is the other state of the same pair, p denotes the pair of i whereas p' denotes the other pair.

$$\begin{aligned} (e^{\mathbf{Q}t})_{ii} &= \pi_i \;+\; \pi_i\frac{\pi_{p'}}{\pi_p}\, e^{-t} \;+\; \frac{\pi_{i'}}{\pi_p}\, e^{-(\kappa_p\pi_p+\pi_{p'})t} \\ (e^{\mathbf{Q}t})_{ii'} &= \pi_{i'} \;+\; \pi_{i'}\frac{\pi_{p'}}{\pi_p}\, e^{-t} \;-\; \frac{\pi_{i'}}{\pi_{p'}}\, e^{-(\kappa_p\pi_p+\pi_{p'})t} \\ (e^{\mathbf{Q}t})_{ij} &= \pi_j\,\left(1 \;-\; e^{-t}\right) \qquad\qquad j \notin p \;. \end{aligned} \tag{2.20}$$

Kimura-80 matrices. Kimura-80 (in short, K80, or sometimes K2P, see [13]) are matrices of the form TN93 with $\pi_A = \pi_C = \pi_G = \pi_T = 1/4$ and $\kappa_R = \kappa_Y = \kappa$.

The transition probabilities of Equation (2.20) read as:

$$\begin{aligned}
(e^{\mathbf{Q}t})_{ii} &= \frac{1}{4}\left(1 + e^{-t} + 2e^{-t(\kappa+1)/2}\right) \\
(e^{\mathbf{Q}t})_{ii'} &= \frac{1}{4}\left(1 + e^{-t} - 2e^{-t(\kappa+1)/2}\right) \\
(e^{\mathbf{Q}t})_{ij} &= \frac{1}{4}\left(1 - e^{-t}\right) \qquad\qquad j \neq i, j \neq i' \ .
\end{aligned} \tag{2.21}$$

2.1.2.6 Some Birth-Death Processes

The explicit diagonalization of the generator is known for some birth and death processes.

The Constant-Rate Birth-Death Process. The constant-rate birth-death process was studied in Section 2.1.1.3. We consider here the truncated version with N states, see also Section 2.1.1.7. The generator of the process is the matrix of $\mathcal{M}_{N\times N}$:

$$\mathbf{M}(N;\lambda,\mu) = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\mu+\lambda) & \lambda & & \\ & \ddots & \ddots & \ddots & \\ & & \mu & -(\mu+\lambda) & \lambda \\ & & & \mu & -\mu \end{pmatrix} . \tag{2.22}$$

The eigenvalues of this matrix are 0, and

$$\omega_k = -(\lambda+\mu) + 2\sqrt{\lambda\mu}\cos\frac{k\pi}{N} , \qquad k = 1..N-1 .$$

Let $\rho = \lambda/\mu$. For $k, p \in \{1, \ldots, N-1\}$, define the values:

$$\begin{aligned}
\phi_{pk} &= \sin\frac{pk\pi}{N} - \rho^{-1/2}\sin\frac{p(k-1)\pi}{N} \\
\gamma_k &= \frac{N}{2}\left(1 - 2\rho^{-1/2}\cos\frac{k\pi}{N} + \rho^{-1}\right) .
\end{aligned}$$

For $k = 0$, define:

$$\begin{aligned}
\phi_{p0} &= \rho^{(p-1)/2} \\
\gamma_0 &= \frac{1-\rho^N}{1-\rho} ,
\end{aligned}$$

if $\rho \neq 1$, and $\gamma_0 = N$ if $\rho = 1$. Finally, define the matrices:

$$\mathbf{\Phi} = ((\phi_{pk}))_{1\leq p\leq N, 0\leq k<N}, \qquad \mathbf{\Psi} = \mathbf{\Phi}' = ((\phi_{kp}))_{0\leq k<N, 1\leq p\leq N} .$$

$$\mathbf{\Omega} = \mathrm{diag}(\omega_0, \ldots, \omega_N) \qquad \mathbf{\Gamma} = \mathrm{diag}(\gamma_0, \ldots, \gamma_N) .$$

We finally can state the result:

Lemma 2.3. *The matrix* $\mathbf{M}(N;\lambda,\mu)$ *is diagonalizable if and only if* $0 \neq \lambda\mu$. *In that case,*

$$\mathbf{M}(N;\lambda,\mu) \;=\; \mathbf{\Gamma}^{-1/2}\mathbf{\Psi}\mathbf{\Omega}\mathbf{\Phi}\mathbf{\Gamma}^{-1/2} \;. \tag{2.23}$$

Using Lemma 2.3, it is possible to write down explicit formulas for the transition probabilities $e^{\mathbf{M}t}$ and the transient probabilities.

As an illustration, consider the case $N = 4$. We have then, defining $\sigma = \rho^{-1/2}$ to simplify expressions:

$$\begin{aligned}
\mathbf{\Phi} &= \begin{bmatrix} 1 & \sqrt{1/2} & 1 & \sqrt{1/2} \\ \sigma^{-1} & \sqrt{2}-\sigma\sqrt{1/2} & -\sigma & -\sqrt{2}-\sigma\sqrt{1/2} \\ \sigma^{-2} & \sqrt{1/2}-\sigma\sqrt{2} & -1 & \sqrt{1/2}+\sigma\sqrt{2} \\ \sigma^{-3} & -\sigma\sqrt{1/2} & \sigma & -\sigma\sqrt{1/2} \end{bmatrix} \\
\mathbf{\Gamma} &= \mathrm{diag}\left(\frac{1-\rho}{1-\rho^4}, 4(1-\sqrt{2}\sigma+\sigma^2), 4(1+\sigma), 4(1+\sqrt{2}\sigma+\sigma^2)\right) \\
\mathbf{\Omega} &= \mathrm{diag}\left(0, -(\lambda+\mu)+\sqrt{2\lambda\mu}, -(\lambda+\mu), -(\lambda+\mu)-\sqrt{2\lambda\mu}\right) \;.
\end{aligned}$$

A Variable-Rate Birth-Death Process. Consider now the following generator $\mathbf{Q} = \mathbf{M}(N;\lambda,\mu)$, where $N \in \mathbb{N}$, λ and μ are strictly positive real numbers:

$$\mathbf{M}(N;\lambda,\mu) \;=\; \begin{pmatrix} -N\lambda & N\lambda & & & & \\ \mu & -(\mu+(N-1)\lambda) & (N-1)\lambda & & & \\ & \ddots & \ddots & \ddots & & \\ & & (N-1)\mu & -((N-1)\mu+\lambda) & \lambda & \\ & & & N\mu & -N\mu & \end{pmatrix} . \tag{2.24}$$

This is a Birth and Death process with transition rates

$$\lambda_k = (N-k)\lambda,\ 0 \le k \le N \qquad \mu_k = k\mu,\ 0 \le k \le N$$

and 0 otherwise. The transition diagram corresponding to this generator is represented in Figure 2.8. The stationary distribution of this process, is according to Expressions (2.7)– (2.8), is the **binomial distribution**

$$\pi(n) \;=\; \frac{1}{(\lambda+\mu)^N}\binom{N}{n}\lambda^n\mu^{N-n} \;.$$

This process appears in many situations that can be interpreted as the superposition of several identical, two-state "particles". One example is the absorption/emission process in lasers: the state of the system is then the number of particles in the "excited" state. Another instance occurs in the modeling of sources of traffic in telecommunications: the state of the system is then the number of sources that are currently active.

Define, for $k \in \{0,\ldots,N\}$:

$$\omega_k \;=\; -k(\lambda+\mu) \;. \tag{2.25}$$

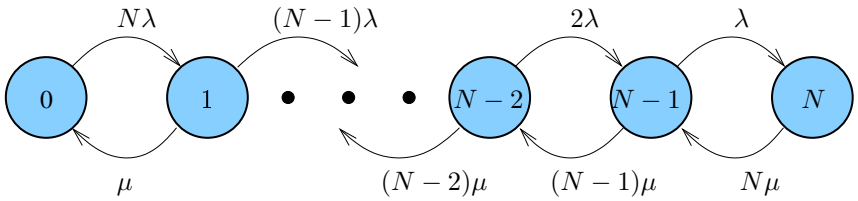

Figure 2.8: A Birth and Death Process

Define further: $\sigma = -\mu/\lambda$. The numbers 1 and σ are the roots of the equation

$$\lambda\sigma^2 \; - \; (\lambda - \mu)\sigma \; - \; \mu \; = \; 0. \tag{2.26}$$

Let $\boldsymbol{\phi}_k = (\phi_k(0), \ldots, \phi_k(N))$ be the row vector with coordinates given by the coefficient of x^i in the polynomial function $(x-1)^k(x-\sigma)^{N-k}$:

$$\begin{aligned} \phi_k(i) \;\; &= \;\; [x^i](x-1)^k\,(x-\sigma)^{N-k} \\ &= \;\; (-1)^i \sum_{j=0}^{i} \binom{k}{j}\binom{N-k}{i-j}\sigma^{i-j} \; . \end{aligned} \tag{2.27}$$

Finally, define the $(N+1)\times(N+1)$ matrices out of the row vectors $\boldsymbol{\phi}_i$ and the *column* vectors $\boldsymbol{\psi}_j$:

$$\boldsymbol{\Phi} \;\; = \;\; ((\; \phi_i(j)))_{i,j}, \qquad \boldsymbol{\Psi} \;\; = \;\; ((\; \psi_j(i)))_{i,j} \;\; := \;\; ((\; \phi_i(j)\; \sigma^{i+j-N}))_{i,j}. \tag{2.28}$$

We then have the following result (see [2, 19, 1]). For this algebraic result, we consider that λ and μ are arbitrary complex numbers, although in the Markov chain application, these are positive real numbers.

Lemma 2.4. *The matrix* $\mathbf{M}(N;\lambda,\mu)$ *is diagonalizable if and only if* $0 \neq \lambda + \mu$. *In that case, if* $\boldsymbol{\Omega} = diag(\omega_0, \ldots, \omega_N)$,

$$\mathbf{M}(N;\lambda,\mu) \;\; = \;\; (\sigma-1)^{-N} \;\; \boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Phi} \; . \tag{2.29}$$

The fact that $\boldsymbol{\Omega}\boldsymbol{\Phi} = \boldsymbol{\Phi}\mathbf{M}$ is a direct application of the analysis of [2] and [19].
In order to establish (2.29), it remains to be proved that:

$$\boldsymbol{\Phi}\;\boldsymbol{\Psi} \;\; = \;\; (\sigma-1)^N \; \mathbf{I} \; . \tag{2.30}$$

By definition, we have:

$$\begin{aligned}
\sum_{k=0}^{N} \phi_{ik}\psi_{kj} &= \sum_{k=0}^{N} [x^k](x-1)^i(x-\sigma)^{N-i}\phi_{kj}\sigma^{k+j-N} \\
&= \sigma^{j-N} \sum_{k=0}^{N} [x^k](x-1)^i(x-\sigma)^{N-i}[y^j](y-1)^k(y-\sigma)^{N-k}\sigma^k \\
&= \sigma^{j-N}[y^j](y-\sigma)^N \ (x-1)^i(x-\sigma)\big|_{x=\sigma(y-1)/(y-\sigma)} \\
&= \sigma^{j-N}[y^j](y-\sigma)^N \frac{(\sigma(y-1)-1(y-\sigma))^i\sigma^{N-i}(y-1-y+\sigma)^{N-i}}{(y-\sigma)^N} \\
&= \sigma^{j-i}(\sigma-1)^N[y^j]y^i \\
&= (\sigma-1)^N \ \delta_{i=j} \ .
\end{aligned}$$

From (2.30), we can conclude (2.29) provided that $1-\sigma \neq 0$, or equivalently, $\lambda+\mu=0$. This is the condition required in Lemma 2.4.

Using Lemma 2.4, it is possible to write down explicit formulas for the transition probabilities $e^{\mathbf{M}t}$ and the transient probabilities.

As an illustration, consider the case $N=3$. We have then:

$$\begin{aligned}
\mathbf{M}(3;\lambda,\mu) &= (\sigma-1)^{-3} \ \mathbf{\Psi\Omega\Phi} \\
&= \frac{1}{(\sigma-1)^3} \begin{bmatrix} -1 & 3 & -3 & 1 \\ -1 & 2+\sigma & -1-2\sigma & \sigma \\ -1 & 1+2\sigma & -(\sigma+2)\sigma & \sigma^2 \\ -1 & 3\sigma & -3\sigma^2 & \sigma^3 \end{bmatrix} \mathbf{\Omega} \begin{bmatrix} -\sigma^3 & 3\sigma^2 & -3\sigma & 1 \\ -\sigma^2 & 2\sigma+\sigma^2 & -1-2\sigma & 1 \\ -\sigma & 1+2\sigma & -\sigma-2 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix}
\end{aligned}$$

where

$$\mathbf{\Omega} = -(\lambda+\mu)\mathrm{diag}\,(0,1,2,3) \ .$$

2.2 Numerical computations

This section is devoted to methods that compute the probabilities of interest, in a numerical way, with little or no attempts at exploiting the structure of the chain, as it was the case in Section 2.1.

There are two types of questions that one should be concerned with, when considering numerical algorithms:

numerical precision when computers execute arithmetic operations, they typically introduce (small) errors. Some errors are due to the fact that numbers are stored in a memory with limited precision. This causes "roundoff" errors. Likewise, operations such as multiplication or division are executed only approximately. When several operations are chained, the errors tend to propagate and are amplified: this phenomenon is called "numerical unstability".

A consequence of this is that some operations are considered as undesirable because they tend to produce errors: adding or subtracting numbers with very different orders of magnitude, and manipulating numbers of different signs.

Another source of errors lies in the computation of some non-elementary functions such as `exp()`, `sin()`, `log()` etc.

algorithmic complexity there exist usually several methods (or algorithms) to compute a certain result. Some may use more time and/or memory than others. A criterion used for comparing algorithms is their *algorithmic complexity*. Evaluating the complexity of a given algorithm consists in computing the number of "elementary" operations it performs. The exact interpretation of what is elementary depends on the context. Elementary operations may be arithmetic operations (`+`, `-`, `*`, `/`), binary operations, assigments of values to the memory, or calls to some function.

Very often, this number of operations does not depend on the exact value of the data, but on some "size" parameter. For instance, numerical algebra algorithms typically have a complexity that depends on the size n of the matrix, and not on the precise entries of the matrix. Likewise, graph algorithms have typically complexities that depend on the number of nodes n, of edges m or some other graph parameter, and not on the exact configuration of nodes and edges.

Even when the number of operations of the algorithm does depend on the data, it is often possible to talk about the *worst case* complexity. Evaluating it provides a bound on the running time.

In some cases, it is possible to compute exactly the number of operations as a function of n (or other parameters). However, the most useful information is the behavior of this count when n is large. Accordingly, the presentation of the complexity is usually limited to the highest order term, as in $2n^2/3$, $7n^3/8$ for instance. It is often sufficient to talk about the order of magnitude of the complexity. One uses then the "big-oh" notation as in $O(n^2)$ or $O(n^3)$.

In summary, the choice of a numerical algorithm is usually governed by a compromise between both aspects: numerical accuracy and speed. It is worthless to use a fast algorithm which does not compute the right answer! On the other hand, performing exact, full precision calculations, is usually too time consuming. Specialists of numerical computations (see *e.g.*[9]) have selected families of methods which have a reasonable numerical accuracy and a reasonable algorithmic complexity.

In the following, we review some of the principal candidate methods for computing stationary, then transient probabilities for Markov chains.

2.2.1 General numerical methods for stationary probabilities

Recall from the beginning of Section 2 that computing the stationary probability of some Markov chain amounts to solving the linear system $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ or $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, with the additional constraint that $\boldsymbol{\pi}.\mathbf{1} = 1$.

2.2.1.1 Direct solution of the linear system

The direct method consists in simply solving the linear system. In both the DTMC and CTMC cases, the system to be solved is of the form:

$$\boldsymbol{\pi}\mathbf{A} \;=\; \mathbf{0}\;, \qquad \boldsymbol{\pi}.\mathbf{1} = 1$$

with either $\mathbf{A} = \mathbf{P} - \mathbf{I}$ or $\mathbf{A} = \mathbf{Q}$. If $\mathbf{P}$ or $\mathbf{Q}$ is irreducible , then the rank of $\mathbf{A}$ is $n-1$ (if n denotes the size of the matrix, which is also the cardinal of $\mathcal{E}$). This means that any row or column is "redundant" in the linear system, in the sense that it is a linear combination of the other rows or columns.

Let us denote with $\overline{\mathbf{A}}$ the $n \times (n-1)$ matrix obtained from $\mathbf{A}$ by deleting the n-th column (or actually any column). The linear system to be solved is equivalent to the new one:

$$\left(\begin{array}{c} \overline{\mathbf{A}}^t \\ \hline \mathbf{1} \end{array} \right) . \boldsymbol{\pi}^t = \left(\begin{array}{c} 0 \\ \vdots \\ \hline 0 \\ 1 \end{array} \right) .$$

Exceptionnally, we have written this linear equation in the usual "matrix times vector" form, since this is the convention used by most software libraries which provide functions for numerically solving linear systems. When using them, never forget to transpose the matrix $\mathbf{P} - \mathbf{I}$ or $\mathbf{Q}$ and be careful not to truncate a column of $\mathbf{A}$ instead of a row.

The numerical accuracy of this method is not considered as good, because the operation is equivalent to inverting a matrix, which is not usually very stable. The result is usually satisfying however for matrices of small size: n typically up to 20 or 30.

The algorithmic complexity is that of solving a linear system of equations: of the order of n^3 (arithmetic) operations.

This complexity is considered as relatively large. The direct method is considered as inappropriate for large values of n (above 100, say) both because of problems of accuracy and algorithmic complexity.

2.2.1.2 The power method

The power method for DTMC uses Theorem 1.6, specifically, see (1.15):

$$\lim_{n \to \infty} \boldsymbol{\pi}_0 \, \mathbf{P}^n = \boldsymbol{\pi} .$$

The idea is then to compute $\boldsymbol{\pi}_n$ for successive values of n until some N. If N is "large enough", the vector $\boldsymbol{\pi}_N$ should be "close enough" to $\boldsymbol{\pi}$. Note that convergence occurs only if $\mathbf{P}$ is aperiodic. This idea will not work if this is not the case. Note also that convergence occurs theoretically for any choice of the initial vector $\boldsymbol{\pi}_0$. However, in practice, some choices of $\boldsymbol{\pi}_0$ are better than others.

It is important to realize that this method is *approximate*: its aim is not to compute $\boldsymbol{\pi}$ but an approximation of if. Even if numerical computations were perfectly accurate, this algorithm would not give the exact answer.

In practice, it is not easy to determine the value of N such that the vector computed is within a distance ε of the exact value $\boldsymbol{\pi}$. According to Theorem 1.6, the distance between $\boldsymbol{\pi}_n$ and $\boldsymbol{\pi}$ tends to 0 as:

$$||\boldsymbol{\pi}_n - \boldsymbol{\pi}|| = \mathcal{O}(\rho^n) ,$$

which means that there is some constant C such that:

$$||\boldsymbol{\pi}_n - \boldsymbol{\pi}|| \leq C \, \rho^n .$$

for all ρ such that $|\lambda_2| < \rho < 1$, where $|\lambda_2|$ is the second largest eigenvalue of $\mathbf{P}$. So, in order to be sure that the distance is less than some ε, the number n should be such that:

$$C\rho^n \leq \varepsilon \iff n \geq \frac{\log C/\varepsilon}{\log 1/\rho} = \frac{\log C - \log \varepsilon}{\log 1/\rho} .$$

Since neither C nor λ_2 are known exactly in practice, this rule does not provide a concrete way to determine N. It is important however, since it says that the closer $|\lambda_2|$ is to 1, the larger N should be.

In practice, one uses a *stopping rule*. For instance, given a desired accuracy ε, one runs the algorithm until the value N such that:

$$||\boldsymbol{\pi}_N - \boldsymbol{\pi}_{N-1}||_\infty := \max_{s\in\mathcal{E}} |\boldsymbol{\pi}_N(s) - \boldsymbol{\pi}_{N-1}(s)| \leq \varepsilon .$$

This property *does not* imply that each value $\boldsymbol{\pi}_N(s)$ is closer than ε to the exact value $\boldsymbol{\pi}(s)$.

The numerical stability of this iterative algorithm is considered as good, because all numbers involved are positive, and between 0 and 1. There may be problems with orders of magnitude. The usual consequence is that the sum of the entries of vector $\boldsymbol{\pi}_n$ is not exactly 1. In that case, it is necessary to renormalize this vector periodically. This adjustment does not change the convergence behavior.

To summarize, an approximation of the stationary distribution can be computed using Algorithm 1 or one of its variants. Several points need to be made precise for the algorithm to be operational. First, about the initial choice for the vector. Since particular vectors (such as distributions concentrated on a particular state, or uniform distributions) can result in a pathological behavior, some authors advocate the choice of a random vector of probabilities. Second, the period for renormalizations has to be chosen. Renormalization is done using Algorithm 2. The choice may depend on the relative cost of one multiplication and one renormalization, see below.

The algorithmic complexity of Algorithm 1 is N times the number of operations required for a vector/matrix multiplication, where N is the final number of iterations (see above). The cost of one vector/matrix multiplication depends in turn depends on the way the matrix is represented. In the worst case, this is $O(n^2)$ operations. With sparse matrices, this is $O(m)$. See Appendix B.1.2 for details. The cost of a renormalization with Algorithm 2 is $2n$ operations. Depending on the value of m, this may be negligible in front of a multiplication, or be comparable. In the first case, renormalization can occur frequently at little cost.

The power method for CTMC. When the problem is to find $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$, it is possible to use *uniformization*. Indeed, according to Theorem 1.19, the stationary distribution of $\mathbf{Q}$ and any uniformized matrix $\mathbf{P}_\nu = \mathbf{I} + \nu^{-1}\mathbf{Q}$ is the same.

The Power method for DTMC can be used with any value of ν that satisfies the uniformization condition (1.33). However, choosing ν too large has the effect that the eigenvalues of $\mathbf{P}$ get close to 1. As we have seen above, this makes convergence slower. On the other hand, if ν is chosen as the minimum allowed by Condition (1.33), there is a possibility that the matrix $\mathbf{P}_\nu$ be periodic, in which case the algorithm does not converge. The choice of ν should produce a compromise between these two dangers.

Algorithm 1: The power method for computing approximate stationary distributions

Data: A probability transition matrix $\mathbf{P}$
Data: An integer parameter K for periodical renormalizations
Data: A precision parameter ε
Result: A distribution which approximates the stationary distribution of $\mathbf{P}$
begin
  $x \leftarrow$ any probability distribution
  $n \leftarrow 0$
  repeat
    $n \leftarrow n + 1$
    $y \leftarrow x \times \mathbf{P}$
    $dist \leftarrow ||x - y||$
    $x \leftarrow y$
    if $n \ mod \ K = 0$ **then**
      $x \leftarrow$ `Renormalize`(x)
  until $dist < \varepsilon$;
  $x \leftarrow$ `Renormalize`(x)
  return x

Algorithm 2: `Renormalize(x)`: Renormalization of a vector

Data: A vector of numbers x indexed by $\mathcal{E}$
Result: A vector proportional to x which entries sum up to 1
begin
  $sum \leftarrow 0$
  for $i \in \mathcal{E}$ **do**
    $sum \leftarrow sum + x[i]$
  for $i \in \mathcal{E}$ **do**
    $x[i] \leftarrow x[i]/sum$
  return x

2.2.1.3 Relaxation.

The method of *successive over relaxation* has been proposed in numerical linear algebra to, at the same time, speed up the calculation of solutions to linear systems, and reduce memory consumption by storing only one vector.

The principle, applied to DTMCs, is to rewrite the system of equations to be solved as:

$$(1 - P_{ii})\ \pi(i) \ = \ \sum_{j \neq i} \pi(j) P_{ji},$$

or:

$$\pi(i) \ = \ \sum_{j \neq i} \pi(j)\ \frac{P_{ji}}{1 - P_{ii}}, \qquad i \in \mathcal{E}\ .$$

Then, one fixes some $\omega \in (0, 2)$ and iteratively computes a sequence of vectors $\pi^{(1)}, \ldots, \pi^{(k)}, \ldots$ according to the formula:

$$\pi^{(k+1)}(i) \ = \ (1 - \omega)\pi^{(k)}(i) \ + \ \frac{\omega}{1 - P_{ii}} \left(\sum_{j<i} \pi^{(k+1)}(j)\ P_{ji} \ + \ \sum_{j \neq i} \pi^{(k)}(j)\ P_{ji} \right)\ .$$

When this computation is performed with increasing values of i, the special form of this formula allows to store the result $\pi^{(k+1)}$ in the same array as the data $\pi^{(k)}$: after state i is handled, the value of $\pi^{(k)}(i)$ will never be used again. In addition, with a proper choice of ω the convergence may be faster than that of the power method. Unfortunately, there is no simple way to compute the best value, or even a "good" value. The value $\omega = 1.2$ is recommended in [26]. It may even happen that the method is unstable and does not converge at all.

With the choice of $\omega = 1$, this reduces to the method of Gauss-Seidel. This method is known to converge under the sufficient (but not necessary) condition that $P_{ii} \geq \sum_{j \neq i} P_{ij}$ for all $i \in \mathcal{E}$.

2.2.2 General methods for transient probabilities

Several situations may occur in practice. First of all, the computation may involve the probability vector(s) $\boldsymbol{\pi}_t$ or the stochastic matrices $\mathbf{P}^n$ or $e^{t\mathbf{Q}}$. Then, assuming for instance that we are concerned with vectors, the problem may be (with small variations between discrete and continuous time):

A) to compute $\boldsymbol{\pi}_n$ for *all* values of n and some $\boldsymbol{\pi}_0$ (in discrete time), or $\boldsymbol{\pi}_{t_i}$ for a sequence of n time instants t_i (in continuous time):

B) to compute $\boldsymbol{\pi}_n$ for *some* value of n and some $\boldsymbol{\pi}_0$;

C) to compute $\boldsymbol{\pi}_n$ for *some* value of n and *many* values of $\boldsymbol{\pi}_0$;

D) to compute an expected value of the form

$$\sum_{s \in \mathcal{E}} v(s)\pi_n(s)$$

for some given vector $\mathbf{v}$ and some value of n,

and combinations of these. The choice of the best algorithm effectively depends on the problem, the value of n and the size of the state space $\mathcal{E}$. We therefore provide here elements of information to guide the decision.

2.2.2.1 Discrete-time chains

Recurrences. Using the recurrence $\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n \mathbf{P}$ provides the values of $\boldsymbol{\pi}_n$ for all $0 \leq n \leq N$, and so solves Problem A) as well as problem B). This recurrence is actually Algorithm 1, in a variant where the iteration stops after N steps, and not when some convergence condition is encountered. The algorithmic complexity of this is $N \times O(m)$. See the comments above.

In certain cases, it is useful to compute $\mathbf{P}^n$ for some specific value n.

The question of computing a power using a minimal number of multiplications is the topic of [15, Section 4.6.3]. For instance, if $n = 2^\ell$, one can use the *doubling formula*:

$$\mathbf{A}^{2^\ell} \;=\; \left(\mathbf{A}^{2^{\ell-1}}\right)^2 \tag{2.31}$$

which allows to compute the power $\mathbf{P}^n$ in $\ell = \log_2(n)$ matrix multiplications. Using this technique, any power $\mathbf{P}^n$ can be computed in less than $2\log_2(n)$ multiplications.

Diagonalization. Assume that $\mathbf{P}$ has been numerically diagonalized. Since we have the spectral expansion: (A.8) (see Theorem A.2 ii/):

$$\mathbf{P}^n \;=\; \sum_{i=1}^{p} \lambda_i^n \; \mathbf{w}_i.\mathbf{v}_i$$

then the general formula for vector $\boldsymbol{\pi}_n$ is:

$$\boldsymbol{\pi}_n \;=\; \sum_{i=1}^{p} \lambda_i^n \; (\boldsymbol{\pi}_0.\mathbf{w}_i) \; \mathbf{v}_i \; .$$

The n scalar products $\boldsymbol{\pi}_0.\mathbf{w}_i$ can be pre-computed. This costs $p \times O(m)$ operations where m is the number of non-zero entries in $\boldsymbol{\pi}_0$. The evaluation of a power λ_i^n is considered as an elementary operation. Then any evaluation of $\boldsymbol{\pi}_n$ consists in computing the linear combination of p vectors of size p: this costs $O(p^2)$ operations, independently of n.

Unfortunately, the Jordan decomposition (including the diagonal decomposition when it exists) is known to be numerically unstable. This is due to the fact that, in some cases, the matrices $\mathbf{R}$ and $\mathbf{S}$ can have entries with different sign and very different magnitudes. Any algorithm which tries to compute them is prone to numerical errors. This method is therefore *not* recommended.

A more robust method that is recommended in [9] and [24] is based on Schur's decomposition and a "quasi-diagonalisation". This topic is too technical to be developped in this version of the document.

2.2.2.2 Continuous-time chains

As in the case of DTMCs, diagonalizing $\mathbf{Q}$ provides a way to compute numerically transient probabilities. But the method is numerically unstable. The corresponding spectral expansion is (assuming that $\mathbf{Q}$ is diagonalizable):

$$e^{\mathbf{Q}t} \;=\; \sum_{i=1}^{p} e^{\lambda_i t} \; \mathbf{w}_i.\mathbf{v}_i$$

and the general formula for vector $\boldsymbol{\pi}_t$ is:

$$\boldsymbol{\pi}_t \;=\; \sum_{i=1}^{p} e^{\lambda_i t}\;(\boldsymbol{\pi}_0.\mathbf{w}_i)\;\mathbf{v}_i\;.$$

Evaluating the exponential series. One idea is to compute directly the exponential series (see A.10), suitably truncated. The general idea is to find N such that the following approximation holds in some sense:

$$e^{t\mathbf{Q}} \;=\; \sum_{n=0}^{\infty} \frac{t^n}{n!}\,\mathbf{Q}^n \;\approx\; \sum_{n=0}^{N} \frac{t^n}{n!}\,\mathbf{Q}^n\;.$$

It is necessary to quantify more precisely this relationship.

First of all, it is possible to choose a number N in such a way that the "rest" of the series is smaller than some ε. The simplest way to do that is to use the following bound: let z be a positive real number. Then

$$R_N(z) \;:=\; \sum_{n=N+1}^{\infty} \frac{z^n}{n!} \;=\; \sum_{n=0}^{\infty} \frac{z^{n+N+1}}{(n+N+1)!} \;\le\; \sum_{n=0}^{\infty} \frac{z^n z^{N+1}}{n!(N+1)!} \;=\; \frac{z^{N+1}}{(N+1)!}\,e^z\;.$$

The sequence $N \mapsto z^N/N!$ is monotonously decreasing for $N \ge z$. It converges rapidly to 0, so that given some ε and some real z, it is numerically fast to find some $N(\varepsilon, z) \ge z$ such that $R_{N(\varepsilon,z)}(z) \le \varepsilon$.

Next, consider the "rest" of the series defining the exponential of a matrix $\mathbf{A}$:

$$R_N(\mathbf{A}) \;:=\; \sum_{n=N+1}^{\infty} \frac{\mathbf{A}^n}{n!} \;=\; e^{\mathbf{A}} \;-\; \sum_{n=0}^{N} \frac{\mathbf{A}^n}{n!}\;.$$

Let $\|\mathbf{A}\|$ denote a *matrix norm* (see Appendix A.6). Since $\|\mathbf{A}^n\| \le \|\mathbf{A}\|^n$, we have:

$$R_N(\mathbf{A}) \;\le\; R_N(\|\mathbf{A}\|).$$

Therefore: if N is chosen as $N(\varepsilon, \|\mathbf{A}\|)$, then the finite sum

$$S_N(\mathbf{A}) \;:=\; \sum_{n=0}^{N} \frac{\mathbf{A}^n}{n!} \tag{2.32}$$

approximates the exponential $e^{\mathbf{A}}$ with an error less than ε, as measured with the matrix norm $\|\cdot\|$. The truncated series S_N can then be evaluated using N matrix additions and N matrix/matrix multiplications.

Let us consider the application of this principle to the matrix $e^{t\mathbf{Q}}$:

$$e^{t\mathbf{Q}} \;=\; \sum_{n=0}^{\infty} \frac{t^n}{n!}\,\mathbf{Q}^n \;=\; S_N(t\mathbf{Q}) \;+\; R_N(t\mathbf{Q})\;.$$

It is convenient to select the *maximum row sum* norm defined as (see also Appendix A.6):

$$|||\mathbf{A}|||_\infty \;=\; \max_i \sum_j |A_{ij}| \;.$$

From the properties of infinitesimal generators, we have:

$$|||t\mathbf{Q}|||_\infty \;=\; t \max_{i\in\mathcal{E}} \left(\sum_{j\neq i} Q_{ij} + |Q_{ii}| \right) \;=\; t \max_{i\in\mathcal{E}} \left(|Q_{ii}| + |Q_{ii}| \right) \;=\; 2t \max_{i\in\mathcal{E}} |Q_{ii}| \;,$$

which should be used to compute $N(\varepsilon, t\,\|\mathbf{Q}\|)$. However, using this method directly with $\mathbf{Q}$ is considered as bad because this matrix has positive and negative terms. In addition, the magnitude of the elements of $\mathbf{Q}^n$ is not bounded when n is large. As a result, numerical unstabilities may appear when evaluating the sum $S_N(t\mathbf{Q})$: the evaluation of this sum with a good precision is also a problem.

The right approach is to use uniformization to transform the problem. For all ν satisfying Condition (1.33), we have (see the proof of Theorem 1.18 p. 16):

$$e^{t\mathbf{Q}} \;=\; e^{t(-\nu\mathbf{I}+(\mathbf{Q}+\nu\mathbf{I}))} \;=\; e^{-\nu t}\, e^{t(\mathbf{Q}+\nu\mathbf{I})} \;=\; e^{-\nu t} \sum_{n=0}^{\infty} \frac{t^n}{n!}\, (\nu\mathbf{I} \;+\; \mathbf{Q})^n \;. \tag{2.33}$$

Then if an approximation within ε is needed, it is enough to obtain an approximation of $e^{t(\mathbf{Q}+\nu\mathbf{I})}$ within $e^{\nu t}\varepsilon$. In addition, we have:

$$\|t(\mathbf{Q}+\nu\mathbf{I})\| \;=\; t \max_{i\in\mathcal{E}} \left(\sum_{j\neq i} Q_{ij} + \nu + Q_{ii} \right) \;=\; t\nu \;.$$

So if ν is chosen as $\max_{i\in\mathcal{E}} |Q_{ii}|$, the norm of the matrix is smaller and it is possible to choose some $N(\varepsilon)$ smaller than with $\mathbf{Q}$. Finally, since the sum $S_N(t(\mathbf{Q}+\nu\mathbf{I}))$ only involves positive matrices, the method is much more stable.

The approximation technique can be applied to compute the $\boldsymbol{\pi}_t$ instead of the matrix $\mathbf{Q}$. Indeed,

$$\boldsymbol{\pi}_t \;=\; \boldsymbol{\pi}_0\, e^{t\mathbf{Q}} \;=\; \sum_{n=0}^{\infty} \frac{t^n}{n!}\, \boldsymbol{\pi}_0 \mathbf{Q}^n \;=\; e^{-\nu t} \sum_{n=0}^{\infty} \frac{t^n}{n!}\, \boldsymbol{\pi}_0\, (\nu\mathbf{I} \;+\; \mathbf{Q})^n \;. \tag{2.34}$$

The error of an approximation of $\boldsymbol{\pi}_t$ is measured with a *vector norm*. In order to apply the method above, one should select a matrix norm compatible with this vector norm, in the sense that $\|\mathbf{A}x\| \leq \|\mathbf{A}\|\;\|x\|$. In that case, we will have:

$$\|\boldsymbol{\pi}_t - \boldsymbol{\pi}_0 S_N(\mathbf{Q})\| \;\leq\; \|\boldsymbol{\pi}_0\|\;\left\|e^{\mathbf{Q}t} - S_N(\mathbf{Q})\right\|$$

which can be made less than some ε by a suitable choice of N. Likewise,

$$\left\|\boldsymbol{\pi}_t - e^{-\nu t}\, \boldsymbol{\pi}_0 S_N(\nu\mathbf{I} \;+\; \mathbf{Q})\right\| \;\leq\; \|\boldsymbol{\pi}_0\|\;\left\|e^{\mathbf{Q}t} - e^{-\nu t} S_N(\nu\mathbf{I} \;+\; \mathbf{Q})\right\| \;.$$

Evaluating the sum $\boldsymbol{\pi}_0 S_N(\mathbf{A})$ for some matrix $\mathbf{A}$ can be done using vector/matrix products and vector additions: it is therefore faster than computing the matrix $S_N(\mathbf{A})$, then multiplying by the vector $\boldsymbol{\pi}_0$.

The maximum row sum for matrices, used above, is compatible with the *sum norm* for vectors:

$$\|\boldsymbol{\pi}\|_1 \;=\; \sum_{i\in\mathcal{E}} |\pi(i)| \;.$$

This norm is always equal to 1 for probability vectors.

Numerical integration. One method is to exploit the fact that the function $t \mapsto \mathbf{P}_t$ is solution of the differential system (1.27) (see Theorem 1.12 page 12):

$$\frac{\mathrm{d}\mathbf{P}_t}{\mathrm{d}t} \;=\; \mathbf{Q}\mathbf{P}_t \;=\; \mathbf{P}_t\mathbf{Q}$$

with initial condition $\mathbf{P}_0 = \mathbf{I}$. A related method uses the fact that $t \mapsto \boldsymbol{\pi}_t$ is solution of the differential system

$$\frac{\mathrm{d}\boldsymbol{\pi}_t}{\mathrm{d}t} \;=\; \boldsymbol{\pi}_t\,\mathbf{Q} \tag{2.35}$$

with initial condition $\boldsymbol{\pi}_0$. This is a consequence of (1.27). The idea is then to integrate numerically this system, using a dedicated solver.

The choice of the solver is important in order to avoid numerical errors. In order to illustrate this, consider Euler's scheme applied to (2.35). Given an integration step δt, the scheme calculates successively approximations $\widetilde{\boldsymbol{\pi}}_{n\delta t}$ according to the recurrence:

$$\widetilde{\boldsymbol{\pi}}_{(n+1)\delta t} \;=\; \widetilde{\boldsymbol{\pi}}_{n\delta t} \;+\; \delta t\;\widetilde{\boldsymbol{\pi}}_{n\delta t}\;\mathbf{Q}\;.$$

Since these are all linear manipulations, the recurrence can be restated as:

$$\widetilde{\boldsymbol{\pi}}_{n\delta t} \;=\; \widetilde{\boldsymbol{\pi}}_{(n-1)\delta t}\;\left(\mathbf{I}+\delta t\mathbf{Q}\right) \;=\; \widetilde{\boldsymbol{\pi}}_0\;\left(\mathbf{I}+\delta t\mathbf{Q}\right)^n \;. \tag{2.36}$$

The good news is that this form is similar to the recurrence for DTMC. Indeed, the matrix in this recurrence is that of a uniformized matrix $\mathbf{P}_{1/\delta t}$ if $\nu := 1/\delta t$ is large enough to satisfy Condition (1.33), or equivalently if δt is smapp enough. As a consequence, the limit of the recurrence is indeed the same as the limit of $\mathbf{P}_t$ when $t \to \infty$ (see also Theorem 1.15). But this does not mean that every term of the recurrence is close to the transient probabilities $\boldsymbol{\pi}_t$.

Euler's scheme is just the simplest of a family of numerical integration methods that have been proposed. Another widely-used one is Runge-Kutta's method of order 4 (abreviated as RK4). In the present situation, it can be shown that it is equivalent to computing

$$\widetilde{\boldsymbol{\pi}}_{n\delta t} \;=\; \widetilde{\boldsymbol{\pi}}_{(n-1)\delta t}\;\left(\mathbf{I}+\delta t\mathbf{Q}+\frac{1}{2}(\delta t)^2\mathbf{Q}^2+\frac{1}{6}(\delta t)^3\mathbf{Q}^3+\frac{1}{24}(\delta t)^4\mathbf{Q}^4\right)\;. \tag{2.37}$$

From Equations (2.36) and (2.37), we see that both numerical integration methods produce approximations of the probabilities that are based on the approximation of the exact equation:

$$\widetilde{\boldsymbol{\pi}}_{(n+1)\delta t} \;=\; \widetilde{\boldsymbol{\pi}}_{n\delta t}\;e^{\mathbf{Q}\delta t} \;=\; \widetilde{\boldsymbol{\pi}}_{n\delta t}\;S_N(\mathbf{Q}\delta t)$$

where S_N is the finite sum defined in (2.32). The value $N = 1$ corresponds to Euler's method, the value $N = 4$ to RK4. In that sense, it is not really useful to perform numerical integration instead of directly evaluating matrix sums.

We have argued that using the sum $S_N(\delta t\mathbf{Q})$ may raise numerical instabilities. The uniformization technique may also be used in this context. If one performs the change of function $\mathbf{q}_t = e^{-\nu t}\boldsymbol{\pi}_t$ in (2.35), one gets the differential system:

$$\frac{\mathrm{d}\mathbf{q}_t}{\mathrm{d}t} \;=\; \mathbf{q}_t\,(\nu\mathbf{I} + \mathbf{Q})\;.$$

Then numerical integration such as in (2.36) or (2.37) can be used, with $\mathbf{Q}$ replaced by $\nu\mathbf{I} + \mathbf{Q}$, a positive matrix. Matrix/vector products become more stable, but the drawback is that the vector function $\mathbf{q}_t$ is not bounded, and grows without limit. This problem does not happen with $\boldsymbol{\pi}_t$ which remains bounded.

Asymptotic approximations. Depending on the range of interest for the time variable t, approximate formulas can be useful.

If t is small, the definition of the exponential of matrices provides the expansion:

$$e^{t\mathbf{Q}} \;\simeq\; \mathbf{I} \;+\; t\mathbf{Q} \;+\; \frac{t^2}{2}\mathbf{Q}^2 \;+\; \frac{t^3}{3!}\mathbf{Q}^3 \;+\; \ldots$$

This is actually the principle of numerical integration methods: as we have seen, they implicitly use this expansion as an approximation.

Problem A) can be addressed by first computing and storing in memory the matrices $\mathbf{Q}^m$, $m \leq m_0$, or the first vectors $\boldsymbol{\pi}_m = \boldsymbol{\pi}_0\mathbf{Q}^m$. The value of m_0 is chosen to be large enough so that errors are bounded by some specified ε. For each s and each t_i, the value of $\boldsymbol{\pi}_{t_i}(s)$ is obtained using $O(m_0)$ operations.

If t is large, then the spectral expansion predicts that:

$$e^{t\mathbf{Q}} \;=\; \sum_{i=1}^{N} e^{\lambda_i t}\,A_i \;\simeq\; \sum_{i=1}^{i_0} e^{\lambda_i t}\,A_i\;,$$

where i_0 is chosen large enough so that the magnitude of the remaining terms is less than some ε. Here, the eigenvalues of $\mathbf{Q}$ have been sorted in increasing order of absolute value.

Combination of methods. One method uses the fact that, for all integer M and all matrix $\mathbf{A}$,

$$e^{\mathbf{A}} \;=\; \left(e^{\mathbf{A}/M}\right)^M\;.$$

If M is large enough, one may use an asymptotic approximation $S_N(\mathbf{A}/M)$ to approximate the inner term. Then raise the result to the appropriate power. The choice of $M = 2^\ell$ is good, because of the doubling formula (2.31). The accuracy of such an approximation is quantified in the following bound [9, Theorem 10.1]: for every matrix norm, every matrix $\mathbf{A}$, and every integer N and M:

$$|||e^{\mathbf{A}} - (S_N(\mathbf{A}/M))^M||| \;\leq\; \frac{|||\mathbf{A}|||^{M+1}}{M^N(N+1)!}\;e^{|||\mathbf{A}|||}\;.$$

The case $M = 1$ corresponds to the bound on $R_N(\mathbf{A})$ we have used above.