

Markov Processes

Michael Manhart

Contents

1	General Properties of Markov Processes	2
1.1	Discrete Time Processes	4
1.1.1	Classification	4
1.1.2	Steady State	6
1.2	Continuous Time Processes	7
1.2.1	Waiting Times and Discrete Time Embedding	8
1.2.2	Classification	9
1.3	Spectral Expansions	9
1.4	Detailed Balance and Reversibility	10
1.4.1	Self-Adjoint Symmetry and Spectral Decomposition	11
1.4.2	Steady State Computation	14
1.4.3	Detailed Balance in Physical Systems	14
1.5	Entropy and the Approach to Steady State	15
1.6	The Backward Equation and Observables	17
1.7	Martingales	18
1.8	The Deterministic Limit	19
2	Markov Processes on 1D Lattices	20
2.1	General Properties	20
2.1.1	Moments and Deterministic Equation	20
2.1.2	Creation and Annihilator Operators	20
2.1.3	Currents and Steady State	21
2.2	Linear Rates	21
2.2.1	Exact Solution	22
2.2.2	Collective Systems	23
2.3	Spectral Expansion	25
2.4	Examples	26
2.4.1	The Poisson Process	26
2.4.2	Symmetric and Asymmetric Random Walks	27
2.4.3	Quantum Harmonic Oscillator and Radiation Field	28
2.4.4	Noninteracting Gas	28
2.4.5	The Moran Population Process	29
2.4.6	Excited Electrons	30
2.4.7	Two-Molecule Chemical Reaction	31

1 General Properties of Markov Processes

Suppose a system can be described by an element of some state space \mathcal{S} . Let X be a random variable taking values in \mathcal{S} . The state of the system over time will be described by some sequence, $\{X(t_1), X(t_2), X(t_3), \dots\}$. Thus, the state is given by a random function $X(t)$ which maps times to values in \mathcal{S} . This function $X(t)$ is said to be a **stochastic process**.¹

In general, the distribution of this variable at different times is described by a hierarchy of functions:

$$\begin{aligned} &P_1(x_1, t_1) \\ &P_2(x_1, t_1; x_2, t_2) \\ &\vdots \\ &P_n(x_1, t_1; \dots; x_n, t_n) \\ &\vdots \end{aligned} \tag{1.1}$$

The function $P_n(x_1, t_1; \dots; x_n, t_n)$ is the probability (or probability density) that X takes on values x_1, \dots, x_n at times t_1, \dots, t_n , respectively. There is an infinite number of such functions, since in general there may be arbitrarily high correlations of the state of the system at different times.

From this hierarchy we can also generate conditional probabilities of the form:

$$P_{i|j}(x_{j+1}, t_{j+1}; \dots; x_{j+i}, t_{j+i} | x_1, t_1; \dots; x_j, t_j) = \frac{P_{j+i}(x_1, t_1; \dots; x_{j+i}, t_{j+i})}{P_j(x_1, t_1; \dots; x_j, t_j)}. \tag{1.2}$$

These are particularly relevant to **Markov processes**, which are a specific class of stochastic processes with a wide range of applicability to real systems. They are defined as stochastic processes such that

$$P_{i|j}(x_{j+1}, t_{j+1} | x_1, t_1; \dots; x_j, t_j) = P_{1|1}(x_{j+1}, t_{j+1} | x_j, t_j). \tag{1.3}$$

In this sense a Markov process is “memoryless”: the probability of X taking a certain value now depends only one of its previous values, not all of them.²³ Then whole hierarchy is specified by only two functions, $P_1(x, t)$ and $P_{1|1}(x_2, t_2 | x_1, t_1)$, since every P_n can be constructed recursively from these two functions using the Markov property:

$$\begin{aligned} P_n(x_1, t_1; \dots; x_n, t_n) &= P_{1|n-1}(x_n, t_n | x_1, t_1; \dots; x_{n-1}, t_{n-1}) P_{n-1}(x_1, t_1; \dots; x_{n-1}, t_{n-1}) \\ &= P_{1|1}(x_n, t_n | x_{n-1}, t_{n-1}) P_{n-1}(x_1, t_1; \dots; x_{n-1}, t_{n-1}). \end{aligned} \tag{1.4}$$

The two functions P_1 (the single-time probability distribution, which will generally be referred to as “the” probability distribution) and $P_{1|1}$ (the “transition probability”) must satisfy three conditions for the Markov process to be well-defined:

1. The Chapman-Kolmogorov condition:

$$P_{1|1}(x_3, t_3 | x_1, t_1) = \sum_{x_2} P_{1|1}(x_3, t_3 | x_2, t_2) P_{1|1}(x_2, t_2 | x_1, t_1). \tag{1.5}$$

¹See [1] for a more detailed discussion on the physical setting of stochastic processes.

²Note the previous observation on which one condition does not have to be immediately previous one. Conditioning on any previous observation fully determines the system. That is, $P_{1|j}(x_{j+1}, t_{j+1} | x_1, t_1; \dots; x_j, t_j) = P_{1|1}(x_{j+1}, t_{j+1} | x_i, t_i)$ for all $i = 1, \dots, j$.

³The condition of being Markovian depends on the state space \mathcal{S} under consideration. For instance, a random walk with constant probability of stepping right or left is clearly Markovian when the state space is simply the current position. However, a random walk with persistence – for example, one in which the probability of stepping right or left depends on the previous step – is *not* Markovian when the state space is the current position, but it *is* Markovian if the state space is enlarged to include both the current position and the previous position. In principle, such a procedure can be implemented to embed many non-Markovian processes into a Markovian process with a bigger state space. However, this is often not useful in practice.

2. The time-evolution equation:

$$P_1(x_2, t_2) = \sum_{x_1} P_{1|1}(x_2, t_2|x_1, t_1) P_1(x_1, t_1). \quad (1.6)$$

3. And finally two normalization conditions:

$$1 = \sum_x P_1(x, t), \quad 1 = \sum_{x_2} P_{1|1}(x_2, t_2|x_1, t_1). \quad (1.7)$$

Two functions satisfying these properties are sufficient to define a Markov process.

Now we introduce two important definitions. A **homogeneous** process is one such that $P_{1|1}(x_2, t_2; x_1, t_1)$ depends only on the time interval $t_2 - t_1$. A **stationary** process is one with homogeneous $P_{1|1}$ and time-independent P_1 . Therefore a stationary process describes systems in steady state: the probability of observing it in any given state does not change with time. It is often possible to embed a merely homogeneous process as a subensemble of a stationary process (and so the stationary process is approached as the steady state). However, this is not always the case. [More details on this, with Wiener process as counterexample.]

If the process is homogeneous, it is especially convenient to formulate the model in the language of linear algebra. We will use the bra-ket notation of quantum mechanics. Define a vector $|x\rangle$ for each state $x \in \mathcal{S}$, and define a vector space \mathcal{V} as the span of all $|x\rangle$. The distribution $P_1(x, t)$ will be represented as components $\pi(x, t)$ of the ket $|\pi(t)\rangle$ in \mathcal{V} with an expansion in the physical basis of states:

$$|\pi(t)\rangle = \sum_x \pi(x, t)|x\rangle. \quad (1.8)$$

We will represent the transition probability $P_{1|1}$ as matrix elements of an operator $\mathbf{\Pi}(t)$: $P_{1|1}(x_2, t_2|x_1, t_1) = \langle x_2|\mathbf{\Pi}(t_2 - t_1)|x_1\rangle$. (Here the homogeneity assumption is essential.) Hence the three conditions from before can be recast in this notation:

1. The Chapman-Kolmogorov condition:

$$\langle x_3|\mathbf{\Pi}(t_3 - t_1)|x_1\rangle = \sum_{x_2} \langle x_3|\mathbf{\Pi}(t_3 - t_2)|x_2\rangle \langle x_2|\mathbf{\Pi}(t_2 - t_1)|x_1\rangle. \quad (1.9)$$

2. The time-evolution equation:

$$\begin{aligned} \pi(x_2, t_2) &= \sum_{x_1} \langle x_2|\mathbf{\Pi}(t_2 - t_1)|x_1\rangle \pi(x_1, t_1) \\ \langle x_2|\pi(t_2)\rangle &= \sum_{x_1} \langle x_2|\mathbf{\Pi}(t_2 - t_1)|x_1\rangle \langle x_1|\pi(t_1)\rangle \\ \longrightarrow \quad |\pi(t_2)\rangle &= \mathbf{\Pi}(t_2 - t_1)|\pi(t_1)\rangle. \end{aligned} \quad (1.10)$$

3. And finally two normalization conditions:

$$1 = \sum_x \pi(x, t), \quad 1 = \sum_{x_2} \langle x_2|\mathbf{\Pi}(t)|x_1\rangle. \quad (1.11)$$

Here the Chapman-Kolmogorov condition clearly implies that $\mathbf{\Pi}(t_1)\mathbf{\Pi}(t_2) = \mathbf{\Pi}(t_1 + t_2)$, which means that $\mathbf{\Pi}(t) = \mathbf{P}^{t/\tau}$ for some constant operator \mathbf{P} [2]. (The time scale τ is required for dimensional reasons.)

The set of physically normalized states $|\pi(t)\rangle$ neither span the whole space \mathcal{V} nor do they constitute a subspace, since the normalization condition does not allow for arbitrary scaling and linear combinations. Note

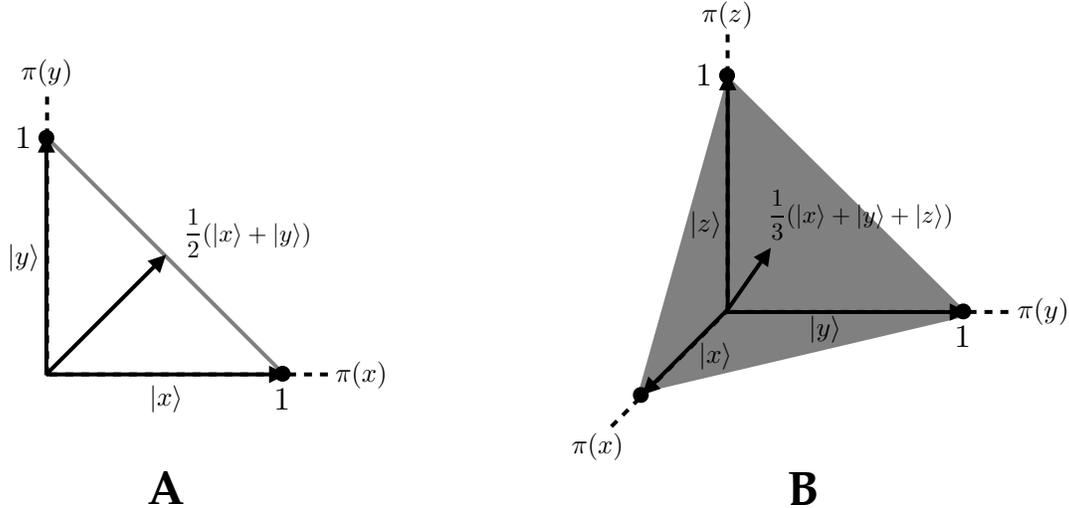


Figure 1: Examples of the hyperplane defining the set of physical states $|\pi(t)\rangle$ along with the state vectors and uniform distribution. (A) $N = 2$, so the physical states constitute a line. (B) $N = 3$, so the states constitute a plane.

that changes of basis will in general lead to coefficients that do not constitute a real probability distribution, so the physical basis is special. The only change of basis preserving this property is a matrix \mathbf{U} such that $\sum_x \langle x | \mathbf{U} | y \rangle = 1$, which is of course the same requirement satisfied by the time evolution matrix $\mathbf{\Pi}(t)$. Thus time evolution can be thought of as a probability-preserving change of basis, although we sometimes also changes of basis in other contexts like permuting the states. The relationship between these pictures is the same as the Schrödinger and Heisenberg pictures of quantum mechanics and will be discussed further in Section 1.6.

In the case of a finite and discrete state space, the set of physical states is easy to visualize. Suppose there are N total states in \mathcal{S} . Then $\mathcal{V} = \mathbb{R}^N$, and the probability vectors $|\pi(t)\rangle$ are confined to an $(N - 1)$ -dimensional hyperplane in the nonnegative quadrant. See Figure 1 for examples with $N = 2, 3$.

1.1 Discrete Time Processes

Since $\mathbf{\Pi}(t) = \mathbf{P}^{t/\tau}$, let us assume there is some smallest time scale Δt so that $t = m\Delta t$ for some integer m and $\mathbf{\Pi}(t) = \mathbf{P}^m$ for some constant single-step transition matrix \mathbf{P} . Then $|\pi(t)\rangle$ can be evolved forward in time in increments of Δt by repeated application of the \mathbf{P} operator:

$$|\pi(t + m\Delta t)\rangle = \mathbf{P}^m |\pi(t)\rangle. \quad (1.12)$$

Thus the matrix elements $\langle x_2 | \mathbf{P} | x_1 \rangle$ give the transition probability from x_1 to x_2 in time step Δt . The limit of infinitesimal Δt will be considered in the next section on continuous time processes.

1.1.1 Classification

Our first goal is to introduce the necessary definitions to classify discrete time processes and characterize their steady states. First we define a set of equivalence classes for states in \mathcal{S} .

Communication classes: We say that state y is **accessible** from state x (denoted $x \rightarrow y$) if it is possible to reach state y from x in a finite number of steps. Mathematically, $\langle y | \mathbf{\Pi}(t) | x \rangle > 0$ for finite t . If both $x \rightarrow y$ and $y \rightarrow x$ (denoted $x \leftrightarrow y$), then x and y fall within the same **communication class** of states. It can be shown that \leftrightarrow defines an equivalence relation on the state space \mathcal{S} , so we can break down the entire space

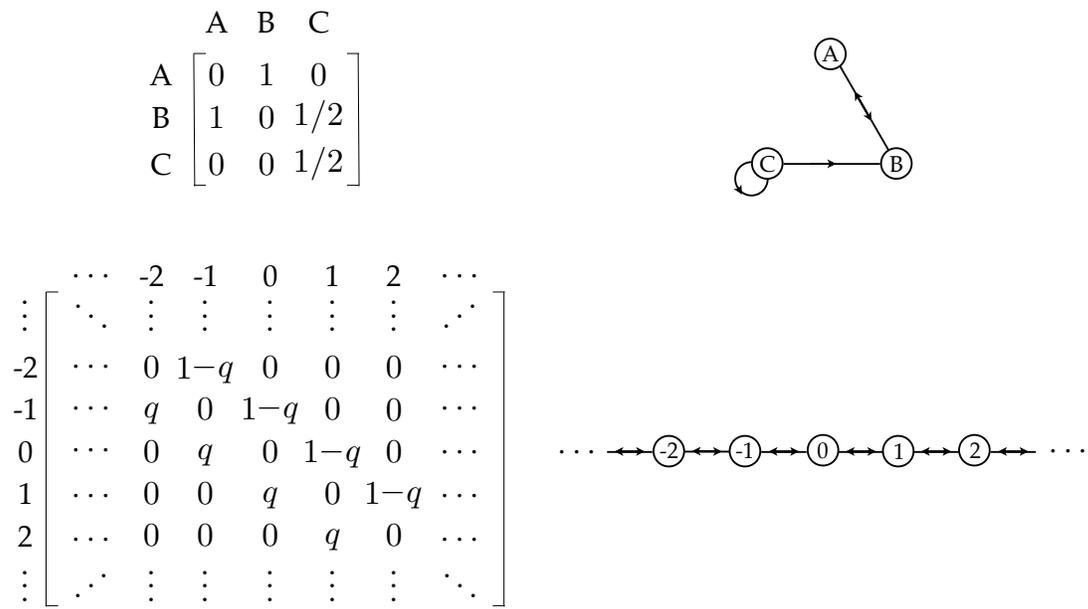


Figure 2: Examples of directed graph representations of Markov processes. On top is a transition matrix \mathbf{P} with its corresponding graph. Here there are two communication classes, $\{A, B\}$ and $\{C\}$. The class $\{A, B\}$ is closed, and the state C is sticky. It is not strongly connected, since it is not possible to access C from A . On the bottom is a transition matrix and graph for the 1D random walk. This graph is strongly connected.

into a set of non-overlapping communication classes. If all \mathcal{S} constitutes a single class, then the process is **irreducible**; otherwise, it is **reducible**. A class may also be **closed** (or **absorbing**) if it is impossible to reach a state outside the class once inside of it. The key feature of communication classes is that important properties of individual states in each class will be shared by *all* states in that class – therefore, our job of classifying processes boils down to classifying communication classes.

When the state space \mathcal{S} is discrete, it is often convenient to visualize properties of the Markov process using a directed graph representation. Nodes are drawn for each state in \mathcal{S} , and for each pair of states x and y , we connect their nodes by an arrow pointing from x to y if $\langle y|\mathbf{P}|x\rangle > 0$ and an arrow from y to x if $\langle x|\mathbf{P}|y\rangle > 0$. Nodes for which $\langle x|\mathbf{P}|x\rangle > 0$ will also have arrows connecting to themselves. Informally, we say such states are “sticky,” since it is possible to stick in that state for successive time steps. Therefore two states are in the same communication class if there are arrows between them in both directions. Graphs are said to be **strongly connected** if one can reach any x from any y in a finite number of steps. This is equivalent to the process being irreducible, since it means that all states are in a single communication class. See Figure 2 for examples.

Periodicity: The period of state x , denoted $d(x)$, is the greatest common divisor of all integers m for which $\langle x|\mathbf{P}^m|x\rangle > 0$. Therefore all closed paths containing x must have lengths that are multiples of $d(x)$. If $d(x) > 1$, then state x is **periodic**. If $d(x) = 1$, then state x is **aperiodic**. If $\langle x|\mathbf{P}^m|x\rangle = 0$ for all m , then define $d(x) = 0$ – it is impossible to return to such a state having left it. Note that periodicity is a property of an *entire* communication class, not just individual states.

If state x is sticky (i.e., $\langle x|\mathbf{P}|x\rangle > 0$), then x must be aperiodic. Thus, if any communication class has a *single* sticky state, then the whole process must be aperiodic. In Figure 2, class $\{A, B\}$ have period 2 in the first example, while $\{C\}$ is aperiodic (it is sticky). For the random walk, each state has period 2. The existence of cyclic structure, however, is not sufficient to guarantee periodicity. It is often more convenient to again think in graph theory terms and consider **cycles** on the graph, which are closed paths. In Figure

Figure 3: Examples of periodicities. (A) This graph clearly has cyclic structure, but because the two cycles have lengths 2 and 3, their GCD is 1 and hence all states are aperiodic. (B) Adding an additional state yields cycles of lengths 2 and 4, which makes the states periodic with period 2.

Figure 4: (A) The Markov process on the left is strongly ergodic in Markovian sense, and this means it also satisfies the ergodic hypothesis of physics.

3A, state C is part of a 2-cycle ($C \rightarrow D \rightarrow C$) and a 3-cycle ($C \rightarrow A \rightarrow B \rightarrow C$), but the GCD of these cycles is 1 since they are coprime. Hence state C is aperiodic, and therefore all states also are since the graph is strongly connected. Contrast this with Figure 3B, in which D is part of a 2-cycle ($D \rightarrow E \rightarrow D$) and a 4-cycle ($D \rightarrow A \rightarrow B \rightarrow C \rightarrow D$). These cycles have GCD 2, so all states are periodic with period 2. Thus to determine periodicity, one must consider the set of all cycles on each class of states.

Transience and recurrence: A state x is **transient** if there is nonzero probability that the system will never return to state x having started there. A state is **recurrent** if the system is guaranteed to return eventually. In that case, the state is **positive recurrent** if the mean recurrence time is finite and **null recurrent** if it is infinite. Again, recurrence and transience are class properties. Note that a recurrent class must be closed. In Figure 2, states A and B are recurrent while state C is transient, since any step away from C prevents future return. These properties will be discussed quantitative detail with more examples in Section ??.

It can be shown that the above classifications are sufficient to characterize Markov processes for many purposes, especially for determining their steady state properties. Thus a process can be classified as follows:

1. Determine the number of communication classes.
2. Characterize each class as either periodic or aperiodic, and as transient, null recurrent, or positive recurrent.

If a class is both aperiodic and positive recurrent, we say it is **strongly ergodic**. If it is aperiodic and null recurrent, it is **weakly ergodic**. This definition of ergodic is related to the standard “ergodic hypothesis” of statistical mechanics, which says that the time average of a system’s observables is equivalent to the average over the statistical ensemble in steady state. This will indeed be true for a Markov process in steady state that is strongly ergodic (as defined here), but it will not be true for a process that is not non-ergodic. For example, in Figure 4A, this process is strongly ergodic. So in steady state, an ensemble average of some observable \mathcal{O} would be

$$\sum_x \mathcal{O}(x) \tilde{\pi}(x) = \frac{1}{3}(\mathcal{O}(A) + \mathcal{O}(B) + \mathcal{O}(C)), \quad (1.13)$$

since the steady state distribution is $1/3$ for each state by symmetry. However, if we took the average of \mathcal{O} at many successive time points, we would still expect this average to be $\frac{1}{3}(\mathcal{O}(A) + \mathcal{O}(B) + \mathcal{O}(C))$, since in steady state we expect the system to spend about $1/3$ of its time in each state.

[Counterexample of something that is non-ergodic in the Markov chain sense, and violates ergodic hypothesis of physics?]

1.1.2 Steady State

Now we will use this scheme to classify steady state behavior. The steady state distribution is defined as

$$|\tilde{\pi}\rangle = \lim_{m \rightarrow \infty} \mathbf{P}^m |\pi(0)\rangle. \quad (1.14)$$

	Periodic	Aperiodic
Transient	Impossible	Impossible
Null recurrent	Impossible	Impossible
Positive recurrent	<i>Steady state:</i> oscillatory <i>Example:</i> circular random walk with odd number of points	<i>Steady state:</i> unique <i>Example:</i> circular random walk with odd number of points

	Periodic	Aperiodic
Transient	<i>Steady state:</i> null <i>Example:</i> 1D asymmetric random walk (no sticking)	<i>Steady state:</i> null <i>Example:</i> 1D asymmetric random walk (with sticking)
Null recurrent	<i>Steady state:</i> null <i>Example:</i> 1D symmetric random walk (no sticking)	<i>Steady state:</i> null <i>Example:</i> 1D symmetric random walk (with sticking)
Positive recurrent	<i>Steady state:</i> oscillatory <i>Example:</i> 1D random walk in potential well (no sticking)	<i>Steady state:</i> unique <i>Example:</i> 1D random walk in potential well (with sticking)

Table 1: Classifications of discrete state space Markov processes, for a finite (top) or infinite (bottom) number of states. The nature of the steady state is given along with an example.

There are three possibilities: the limit exists uniquely, there are multiple limiting distributions (depending on the initial conditions $|\pi(0)\rangle$), or there is no limit at all. We would like to see which possibility is realized for each type of Markov process.

First we consider irreducible processes (only one communication class). There are six possible classifications regarding periodicity and transience/recurrence. We consider only discrete state spaces \mathcal{S} so we may use the language of graph theory developed thus far; continuous state spaces will be treated in a later section. However, we will include both finite and infinite numbers of states. We catalogue the resulting 12 possibilities in Table 1.

What if the process is reducible, i.e., it has multiple communication classes? For a discrete, finite state space, each class must be either positive recurrent or transient and periodic or aperiodic. All transient classes will have zero contribution to the steady state, since all probability will eventually be “absorbed” by the recurrent classes. Each recurrent class will have a linearly independent steady state distribution, and so any properly normalized linear combination of these distributions will be a valid steady state. Hence there is an infinite number of steady states.

1.2 Continuous Time Processes

Now we discuss the continuum limit of time. The evolution equation for a single time step is

$$|\pi(t + \Delta t)\rangle = \mathbf{P}|\pi(t)\rangle. \quad (1.15)$$

Now by subtracting $|\pi(t)\rangle$ from both sides and dividing by Δt we obtain

$$\frac{|\pi(t + \Delta t)\rangle - |\pi(t)\rangle}{\Delta t} = \frac{1}{\Delta t}(\mathbf{P} - \mathbf{1})|\pi(t)\rangle, \quad (1.16)$$

where $\mathbf{1}$ is the identity operator. Taking the limit $\Delta t \rightarrow 0$,

$$\partial_t |\pi(t)\rangle = \mathbf{W}|\pi(t)\rangle, \quad \text{where} \quad \mathbf{W} = \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t}(\mathbf{P} - \mathbf{1}). \quad (1.17)$$

[How is this limit well-defined?] The matrix elements of \mathbf{W} can be expressed specifically as

$$\langle x|\mathbf{W}|y\rangle = \begin{cases} \lim_{\Delta t \rightarrow 0^+} \frac{1}{\Delta t} \langle x|\mathbf{P}|y\rangle, & x \neq y \\ \lim_{\Delta t \rightarrow 0^+} -\frac{1}{\Delta t} \sum_{z \neq x} \langle z|\mathbf{P}|x\rangle = -\sum_{z \neq x} \langle z|\mathbf{W}|x\rangle, & x = y \end{cases} \quad (1.18)$$

since the diagonal elements of \mathbf{P} are $\langle x|\mathbf{P}|x\rangle = 1 - \sum_{y \neq x} \langle y|\mathbf{P}|x\rangle$. Equation 1.17 is known as the **master equation**. If we hit both sides by $\langle x|$ to expand into components, we may write it as

$$\begin{aligned} \partial_t \pi(x, t) &= \sum_y \langle x|\mathbf{W}|y\rangle \pi(y, t) \\ &= \sum_{y \neq x} \langle x|\mathbf{W}|y\rangle \pi(y, t) + \langle x|\mathbf{W}|x\rangle \pi(x, t) \\ &= \sum_{y \neq x} (\langle x|\mathbf{W}|y\rangle \pi(y, t) - \langle y|\mathbf{W}|x\rangle \pi(x, t)). \end{aligned} \quad (1.19)$$

This version of the master equation is perhaps the most illuminating: clearly we can interpret the off-diagonal elements of \mathbf{W} as rates, where the first term on the right-hand side of Equation 1.19 is the probability flux *into* state x , and the second term is the probability flux *out of* state x . Thus the master equation is essentially a continuity equation.

The general solution to the master equation is

$$|\pi(t)\rangle = e^{t\mathbf{W}} |\pi(0)\rangle = \mathbf{\Pi}(t) |\pi(0)\rangle. \quad (1.20)$$

This is obvious from the discrete time case. There the time evolution operator is $\mathbf{\Pi}(t) = \mathbf{P}^{t/\Delta t}$, so

$$\mathbf{\Pi}(t) = \lim_{\Delta t \rightarrow 0^+} \mathbf{P}^{t/\Delta t} = \lim_{\Delta t \rightarrow 0^+} (\mathbf{1} + \Delta t \mathbf{W})^{t/\Delta t} = e^{t\mathbf{W}}. \quad (1.21)$$

The above solution is for a homogeneous process where \mathbf{W} is time-independent. It is however easy to generalize to inhomogeneous cases where the master equation is

$$\partial_t |\pi(t)\rangle = \mathbf{W}(t) |\pi(t)\rangle. \quad (1.22)$$

Then the general solution is

$$|\pi(t)\rangle = \mathcal{T} \{ e^{\int_0^t d\tau \mathbf{W}(\tau)} \} |\pi(0)\rangle, \quad (1.23)$$

where the matrix elements of the time evolution operator are defined as

$$\langle x|\mathcal{T} \{ e^{\int_0^t d\tau \mathbf{W}(\tau)} \} |y\rangle = \langle x|y\rangle + \int_0^t d\tau \langle x|\mathbf{W}(\tau)|y\rangle + \frac{1}{2} \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \langle x|\mathcal{T} \{ \mathbf{W}(\tau_2) \mathbf{W}(\tau_1) \} |y\rangle + \dots \quad (1.24)$$

The notation $\mathcal{T} \{ \dots \}$ is the time-ordering operator. This is required since the rate matrices at different times do not commute in general (i.e., $\mathbf{W}(t_1) \mathbf{W}(t_2) \neq \mathbf{W}(t_2) \mathbf{W}(t_1)$).

1.2.1 Waiting Times and Discrete Time Embedding

Continuous time processes can be embedded in a corresponding discrete time processes. A sample realization of a continuous time process consists of a series of discrete jump events alternating with random waiting times. [Some figure with sample realization of a path.] How are these waiting times distributed? The

probability of remaining in state x after an infinitesimal time step Δt is $1 + \Delta t \langle x | \mathbf{W} | x \rangle$. Suppose we want the probability that the system stays in state x for finite total time $t = m\Delta t$. Then

$$\text{Prob}(X(\tau) = x, \forall \tau \in [0, t] | X(0) = x) = \lim_{m \rightarrow \infty} \left(1 + \frac{t}{m} \langle x | \mathbf{W} | x \rangle \right)^m = e^{t \langle x | \mathbf{W} | x \rangle}. \quad (1.25)$$

Therefore these waiting times are exponentially distributed. Obviously if $\langle x | \mathbf{W} | x \rangle$ is a large negative number, that means there is a large flow out of state x , and hence the waiting times to leave state x will tend to be short; conversely, if $\langle x | \mathbf{W} | x \rangle$ is smaller in magnitude, there is smaller flow out, and hence the waiting times will tend to be longer.

Now we can embed this continuous time process in a discrete time process where the time steps are the waiting times between jumps (even though these are random). We will define the discrete time transition matrix \mathbf{P} as follows. First, diagonal elements must be zero, because by construction a jump occurs at each time step. The only exception is when the state is absorbing in the original process, in which case $\langle x | \mathbf{W} | x \rangle = 0$. Then $\langle x | \mathbf{P} | x \rangle = 1$. So

$$\langle x | \mathbf{P} | x \rangle = \begin{cases} 0, & \text{if } \langle x | \mathbf{W} | x \rangle \neq 0 \\ 1, & \text{if } \langle x | \mathbf{W} | x \rangle = 0 \end{cases} \quad (1.26)$$

Now the off-diagonal elements $\langle x | \mathbf{P} | y \rangle$ must give the probability of transitioning from y to x given that the system leaves state x :

$$\langle x | \mathbf{P} | y \rangle = \lim_{\Delta t \rightarrow 0} \frac{\langle x | \mathbf{\Pi}(\Delta t) | y \rangle}{1 - \langle x | \mathbf{\Pi}(\Delta t) | x \rangle} = - \frac{\langle x | \mathbf{W} | y \rangle}{\langle x | \mathbf{W} | x \rangle}. \quad (1.27)$$

1.2.2 Classification

The classification of continuous time processes now can rely on classification of the embedding in a discrete time process. All the definitions of communication classes, recurrence, and transience carry over, with two exceptions. The first is that while the embedding may exhibit periodicity, it has no direct consequences for the continuous time process in regard to the steady state. Thus we do not discuss it here. Secondly, while a continuous time process is recurrent if and only if the embedding is recurrent, this correspondence does *not* hold for positive and null recurrence. That is, the embedding may be null recurrent even if the continuous time process is positive recurrent. [Example? See Schinazi 1999.] Besides this, the communication classes and the recurrence or transience of a class in a continuous time process are the same as the embedding.

Therefore, for any irreducible, finite state space must be positive recurrent and hence has a unique steady state.

1.3 Spectral Expansions

It is often convenient to analyze spectral properties of the Markov process. Assume the single time-step evolution operator \mathbf{P} (for discrete time processes) has eigenvalues $\lambda_0, \lambda_1, \dots, \lambda_N$ with right and left eigenvectors $|\lambda_j\rangle$ and $\langle \lambda_j|$. (If \mathbf{P} is symmetric, then these eigenvectors are related by a simple transpose. However, in general that will not be true.) For the continuous time rate matrix \mathbf{W} let the spectrum be $\mu_0, \mu_1, \dots, \mu_N$. Then the time evolution operators have the spectral expansions

$$\text{Discrete time: } \mathbf{P}^m = \sum_{j=0}^N \lambda_j^m |\lambda_j\rangle \langle \lambda_j| \quad (1.28)$$

$$\text{Continuous time: } e^{t\mathbf{W}} = \sum_{j=0}^N e^{\mu_j t} |\mu_j\rangle \langle \mu_j|. \quad (1.29)$$

Clearly the leading eigenvalue of \mathbf{P} must be 1 and all other eigenvalues must be smaller than 1; the leading eigenvalue of \mathbf{W} must be 0 and all other eigenvalues must be negative. The left eigenvector with eigenvalue 1 (for \mathbf{P}) or 0 (for \mathbf{W}) is simply the state with 1 as each coefficient:

$$\text{Discrete time: } \langle 1| = \sum_x \langle x| \quad (1.30)$$

$$\text{Continuous time: } \langle 0| = \sum_x \langle x| \quad (1.31)$$

$$(1.32)$$

The right eigenvector with leading eigenvalue will be the steady state $|\tilde{\pi}\rangle$. Hence

$$\text{Discrete time: } \mathbf{P}^m = \sum_x |\tilde{\pi}\rangle \langle x| + \sum_{j=1}^N \lambda_j^m |\lambda_j\rangle \langle \lambda_j| \quad (1.33)$$

$$\text{Continuous time: } e^{t\mathbf{W}} = \sum_x |\tilde{\pi}\rangle \langle x| + \sum_{j=1}^N e^{\mu_j t} |\mu_j\rangle \langle \mu_j|. \quad (1.34)$$

Therefore in the infinite time limit, all the sub-leading eigenvalue terms disappear:

$$\text{Discrete time: } \lim_{m \rightarrow \infty} \mathbf{P}^m = \sum_x |\tilde{\pi}\rangle \langle x| \quad (1.35)$$

$$\text{Continuous time: } \lim_{t \rightarrow \infty} e^{t\mathbf{W}} = \sum_x |\tilde{\pi}\rangle \langle x|. \quad (1.36)$$

So in both cases $\langle y|\mathbf{\Pi}(\infty)|x\rangle = \tilde{\pi}(y)$.

Suppose a left eigenvector with eigenvalue λ has the expansion

$$\langle \lambda| = \sum_x \ell_\lambda(x) \langle x|. \quad (1.37)$$

Then by definition $\langle \lambda|\mathbf{\Pi}(t) = e^{t\lambda}\langle \lambda|$. If we hit both sides by $|x\rangle$ to extract components,

$$\begin{aligned} \langle \lambda|\mathbf{\Pi}(t)|x\rangle &= e^{t\lambda}\langle \lambda|x\rangle \\ \sum_y \ell_\lambda(y) \langle y|\mathbf{\Pi}(t)|x\rangle &= e^{t\lambda}\ell_\lambda(x) \\ \mathbb{E}[\ell_\lambda(X(t))|X(0) = x] &= e^{t\lambda}\ell_\lambda(x). \end{aligned} \quad (1.38)$$

Therefore any function $f(X)$ satisfying $\mathbb{E}[f(X(t))|X(0) = x] = e^{t\lambda}f(x)$ is a left eigenvector with eigenvalues λ .

1.4 Detailed Balance and Reversibility

The coefficients of the steady state distribution $|\tilde{\pi}\rangle = \sum_x \tilde{\pi}(x)|x\rangle$ must satisfy the condition

$$0 = \sum_{y \neq x} (\langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) - \langle y|\mathbf{W}|x\rangle \tilde{\pi}(x)) \quad (1.39)$$

for all $y \in \mathcal{S}$. (In this section, we will focus on continuous time processes. However, these results generally apply to discrete time processes as well.) Clearly it is a sufficient, but not necessary, condition that the function $\tilde{\pi}(x)$ satisfies

$$\langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) = \langle y|\mathbf{W}|x\rangle \tilde{\pi}(x). \quad (1.40)$$

If this condition is met, the system is said to obey **detailed balance**. Note that the quantity $\langle x|\mathbf{W}|y\rangle \pi(y, t)$ describes the probability flux from state y to state x at time t . Hence the detailed balance condition means that in steady state, the fluxes between each pair of state balance. If detailed balance is not obeyed, however, then there is a net current between states even in steady state. Thus there is some cycle in the state space supporting a net current.

Detailed balance is also interpreted as time reversibility in the following sense. Say we have a system in steady state. Then the probability of transitioning from state x to state y in time t is $\langle y|\mathbf{\Pi}(t)|x\rangle \tilde{\pi}(x)$. But detailed balance implies that this equals the probability of the reverse transition, y to x , in time t : $\langle x|\mathbf{\Pi}(t)|y\rangle \tilde{\pi}(y)$. Hence, the dynamics of the system in steady state look the same forward or backward. Obviously this will not be a true in a system without detailed balance; in that case there will be net currents flowing through \mathcal{S} , and forward and backward in time will be distinguishable.

1.4.1 Self-Adjoint Symmetry and Spectral Decomposition

Detailed balance also serves as an important symmetry property of Markov processes. For quantum mechanics and other systems described by linear algebra, the relevant symmetry property of operators is usually Hermiticity or transpose symmetry. For Markov processes, however, this property is too strong. Suppose \mathbf{W} is symmetric. Then any constant function $\tilde{\pi}(x)$ trivially satisfies the steady state condition in Equation 1.39. Thus, a symmetric rate matrix always yields a uniform distribution. In most cases, this is not very interesting.

Detailed balance provides a more useful symmetry property. Besides the fact that it allows for a nontrivial steady state, it guarantees that \mathbf{W} is diagonalizable and hence has a spectral decomposition. First we show this explicitly but less elegantly, and then we show a more elegant argument. We conclude with a specific example for the two state system.

A straightforward procedure is to show how \mathbf{W} is similar to another matrix \mathbf{V} that is transpose symmetric. Hence it is diagonalizable and we can spectrally decompose \mathbf{W} using \mathbf{V} . Assuming \mathbf{W} obeys detailed balance, let us define a symmetric matrix $\mathbf{V} = \mathbf{\Lambda}^{-1}\mathbf{W}\mathbf{\Lambda}$, where $\langle x|\mathbf{\Lambda}|y\rangle = \delta_{xy}\sqrt{\tilde{\pi}(x)}$. So

$$\langle x|\mathbf{V}|y\rangle = \sum_{z,w} \langle x|\mathbf{\Lambda}^{-1}|z\rangle \langle z|\mathbf{W}|w\rangle \langle w|\mathbf{\Lambda}|y\rangle = \sum_{z,w} \frac{\delta_{xz}}{\sqrt{\tilde{\pi}(x)}} \langle z|\mathbf{W}|w\rangle \delta_{wy} \sqrt{\tilde{\pi}(y)} = \sqrt{\frac{\tilde{\pi}(y)}{\tilde{\pi}(x)}} \langle x|\mathbf{W}|y\rangle, \quad (1.41)$$

which can easily be shown to be symmetric. Therefore \mathbf{V} has a real set of eigenvalues $\{\lambda\}$ and a complete basis of eigenvectors $\{|\lambda\rangle_{\mathbf{V}}\}$. The eigenvalues are the same as those of \mathbf{W} , since the two matrices are related by a similarity transformation. The eigenvectors are related in the following manner. Let the right and left eigenvectors of \mathbf{V} be expressed as

$$|\lambda\rangle_{\mathbf{V}} = \sum_x a_{\lambda}(x)|x\rangle, \quad \langle\lambda|_{\mathbf{V}} = \sum_x a_{\lambda}(x)\langle x|. \quad (1.42)$$

The coefficients are the same since they are related by a transpose for symmetric \mathbf{V} . Then it can be shown that the right and left eigenvectors of \mathbf{W} are

$$|\lambda\rangle_{\mathbf{W}} = \sum_x \sqrt{\tilde{\pi}(x)} a_{\lambda}(x)|x\rangle, \quad \langle\lambda|_{\mathbf{W}} = \sum_x \frac{a_{\lambda}(x)}{\sqrt{\tilde{\pi}(x)}} \langle x|. \quad (1.43)$$

Note that the right and left eigenvectors of \mathbf{W} are associated by a factor of the steady state.

Assume the eigenvectors $|\lambda\rangle_{\mathbf{V}}$ are normalized according to the Euclidean inner product $(,)_E$:

$$(|\lambda\rangle_{\mathbf{V}}, |\lambda'\rangle_{\mathbf{V}})_E = \sum_x a_\lambda(x) a_{\lambda'}(x) = \delta_{\lambda, \lambda'}. \quad (1.44)$$

So now we can make the spectral expansion of \mathbf{V} :

$$e^{t\mathbf{V}} = \sum_\lambda e^{\lambda t} |\lambda\rangle_{\mathbf{V}} \langle \lambda|_{\mathbf{V}}. \quad (1.45)$$

The notation $|\lambda\rangle \langle \lambda|$ precisely refers to the outer product of $|\lambda\rangle$ with itself. This is not important here, but will be later. Then since $e^{t\mathbf{W}} = \mathbf{\Lambda} e^{t\mathbf{V}} \mathbf{\Lambda}^{-1}$,

$$\begin{aligned} \langle x | e^{t\mathbf{W}} | y \rangle &= \langle x | \mathbf{\Lambda} e^{t\mathbf{V}} \mathbf{\Lambda}^{-1} | y \rangle = \sum_\lambda e^{\lambda t} \langle x | \mathbf{\Lambda} | \lambda \rangle_{\mathbf{V}} \langle \lambda |_{\mathbf{V}} \mathbf{\Lambda}^{-1} | y \rangle \\ &= \sum_\lambda \sqrt{\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)}}} e^{\lambda t} \langle x | \lambda \rangle_{\mathbf{V}} \langle \lambda |_{\mathbf{V}} y \rangle \\ &= \sum_\lambda \sqrt{\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)}}} a_\lambda(x) a_\lambda(y) e^{\lambda t}. \end{aligned} \quad (1.46)$$

The time-dependent probability distribution can thus be written as

$$\begin{aligned} \pi(x, t) &= \sum_y \langle x | e^{t\mathbf{W}} | y \rangle \pi(y, 0) = \sum_y \pi(y, 0) \sum_\lambda \sqrt{\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)}}} a_\lambda(x) a_\lambda(y) e^{\lambda t} \\ &= \sum_\lambda \sqrt{\tilde{\pi}(x)} a_\lambda(x) e^{\lambda t} \sum_y \frac{a_\lambda(y)}{\sqrt{\tilde{\pi}(y)}} \pi(y, 0). \end{aligned} \quad (1.47)$$

Note that we could have spectrally decomposed \mathbf{W} directly:

$$\begin{aligned} \langle x | e^{t\mathbf{W}} | y \rangle &= \sum_\lambda e^{\lambda t} |\lambda\rangle_{\mathbf{W}} \langle \lambda|_{\mathbf{W}} \\ &= \sum_\lambda e^{\lambda t} \langle x | \left(\sum_z \sqrt{\tilde{\pi}(z)} a_\lambda(z) |z\rangle \right) \left(\sum_w \frac{a_\lambda(w)}{\sqrt{\tilde{\pi}(w)}} \langle w| \right) | y \rangle \\ &= \sum_\lambda \sqrt{\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)}}} a_\lambda(x) a_\lambda(y) e^{\lambda t}. \end{aligned} \quad (1.48)$$

The reason this is more confusing in the above format is that it's not clear how to normalize the eigenvectors of \mathbf{W} – they were not normalized according to the Euclidean inner product. However, it is clear how to normalize the eigenvectors of \mathbf{V} (Euclidean metric), and we can define the eigenvectors of \mathbf{W} with the proper normalization using the eigenvectors of \mathbf{V} .

Now, however, we show a more elegant approach which avoids construction of \mathbf{V} . That is, we can deal with \mathbf{W} directly if we define a new inner product, different from the Euclidean one, under which \mathbf{W} is truly self-adjoint in the mathematical sense. Define the “Markovian”⁴ inner product $(,)_M$ as

$$(|\pi_1(t)\rangle, |\pi_2(t)\rangle)_M \equiv \sum_x \frac{\pi_1(x, t) \pi_2(x, t)}{\tilde{\pi}(x)}. \quad (1.49)$$

⁴This term is my own invention, but the procedure is not.

This satisfies all the conditions of an inner product. Now for this inner product,

$$\begin{aligned}
(|\pi_1(t)\rangle, \mathbf{W}|\pi_2(t)\rangle)_M &= \sum_x \frac{\pi_1(x, t)}{\tilde{\pi}(x)} \sum_y \langle x|\mathbf{W}|y\rangle \pi_2(y, t) \\
&= \sum_{x, y} \frac{\pi_1(x, t)}{\tilde{\pi}(x)} \langle x|\mathbf{W}|y\rangle \pi_2(y, t) \\
&= \sum_{x, y} \frac{\pi_1(x, t)}{\tilde{\pi}(y)} \langle y|\mathbf{W}|x\rangle \pi_2(y, t) \\
&= \sum_y \left(\sum_x \langle y|\mathbf{W}|x\rangle \pi_1(x, t) \right) \frac{\pi_2(y, t)}{\tilde{\pi}(y)} \\
&= (\mathbf{W}|\pi_1(t)\rangle, |\pi_2(t)\rangle)_M.
\end{aligned} \tag{1.50}$$

The third line of the above is where the detailed balance assumption is used. Thus, \mathbf{W} is self-adjoint under the Markovian inner product, and hence we can apply the spectral theorem using this inner product. That is, it has a real set of eigenvalues, and a complete basis of eigenvectors that is orthonormal according to the Markovian inner product. Let the right eigenvectors have the expansion

$$|\lambda\rangle_{\mathbf{W}} = \sum_x b_\lambda(x)|x\rangle. \tag{1.51}$$

Then orthonormality and completeness mean that the coefficients satisfy

$$\sum_x \frac{b_\lambda(x) b_{\lambda'}(x)}{\tilde{\pi}(x)} = \delta_{\lambda, \lambda'}, \quad \sum_\lambda \frac{b_\lambda(x) b_\lambda(y)}{\tilde{\pi}(y)} = \delta_{x, y}. \tag{1.52}$$

Therefore the spectral decomposition is

$$\langle x|e^{t\mathbf{W}}|y\rangle = \sum_\lambda e^{t\lambda} \frac{b_\lambda(x) b_\lambda(y)}{\tilde{\pi}(y)}. \tag{1.53}$$

Identifying $b_\lambda(x) = \sqrt{\tilde{\pi}(x)} a_\lambda(x)$ shows this is equivalent to the previous formulation. Now the time-dependent distribution can be written (in terms of $b_\lambda(x)$)

$$\pi(x, t) = \sum_\lambda b_\lambda(x) e^{\lambda t} \sum_y \frac{b_\lambda(y)}{\tilde{\pi}(y)} \pi(y, 0). \tag{1.54}$$

Since $b_0(x) = \tilde{\pi}(x)$,

$$\pi(x, t) = \tilde{\pi}(x) + \sum_{\lambda < 0} b_\lambda(x) e^{\lambda t} \sum_y \frac{b_\lambda(y)}{\tilde{\pi}(y)} \pi(y, 0). \tag{1.55}$$

As a simple example of this expansion for detailed balance systems, we use the 2×2 case (which trivially obeys detailed balance), with steady state and $\mathbf{\Lambda}$ and \mathbf{V} matrices:

$$\mathbf{W} = \begin{bmatrix} -b & a \\ b & -a \end{bmatrix}, \quad |\tilde{\pi}\rangle = \frac{1}{a+b} \begin{bmatrix} a \\ b \end{bmatrix} \tag{1.56}$$

$$\mathbf{\Lambda} = \frac{1}{\sqrt{a+b}} \begin{bmatrix} \sqrt{a} & 0 \\ 0 & \sqrt{b} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} -b & \sqrt{ab} \\ \sqrt{ab} & -a \end{bmatrix}. \tag{1.57}$$

These coordinates are all in the physical basis. The eigenvalues are 0 and $\mu \equiv -a - b$. Now the eigenvectors of \mathbf{V} are (normalized according to the Euclidean inner product):

$$|0\rangle_{\mathbf{V}} = \frac{1}{\sqrt{a+b}} \begin{bmatrix} \sqrt{a} \\ \sqrt{b} \end{bmatrix}, \quad |\mu\rangle_{\mathbf{V}} = \frac{1}{\sqrt{a+b}} \begin{bmatrix} -\sqrt{b} \\ \sqrt{a} \end{bmatrix}. \quad (1.58)$$

The eigenvectors of \mathbf{W} are (normalized according to the Markovian inner product):

$$|0\rangle_{\mathbf{W}} = \frac{1}{a+b} \begin{bmatrix} a \\ b \end{bmatrix}, \quad |\mu\rangle_{\mathbf{W}} = \frac{\sqrt{ab}}{a+b} \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \quad (1.59)$$

These can readily be plugged into the preceding results to confirm the known result:

$$e^{t\mathbf{W}} = \frac{1}{a+b} \begin{bmatrix} a + be^{-(a+b)t} & a(1 - e^{-(a+b)t}) \\ b(1 - e^{-(a+b)t}) & b + ae^{-(a+b)t} \end{bmatrix}. \quad (1.60)$$

1.4.2 Steady State Computation

Another chief advantage of detailed balance is that it can make computation of the steady state much easier if the rates are known, using Equation 1.40. In general, if there are N discrete states, then Equation 1.40 gives $N(N-1)/2$ equations for each pair of states. However, we only need $N-1$ of these plus the normalization condition to fix the steady state distribution.

A more powerful result can be obtained if the rates $\langle x|\mathbf{W}|y\rangle$ obey certain symmetries. Suppose first that the rates only depend on the difference $x-y$ (translation invariance). Then detailed balance implies

$$\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)} = \frac{\langle x|\mathbf{W}|y\rangle}{\langle y|\mathbf{W}|x\rangle} \equiv \psi(x-y). \quad (1.61)$$

But since

$$\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)} \times \frac{\tilde{\pi}(y)}{\tilde{\pi}(z)} = \frac{\tilde{\pi}(x)}{\tilde{\pi}(z)}, \quad (1.62)$$

this implies that $\psi(x-y)\psi(y-z) = \psi(x-z)$, or $\psi(a)\psi(b) = \psi(a+b)$. This constrains the functional form of ψ to be an exponential function: $\psi(x) = e^{cx}$ for some constant c . Then

$$\frac{\tilde{\pi}(x)}{\tilde{\pi}(y)} = \psi(x-y) = e^{c(x-y)} \quad (1.63)$$

so $\tilde{\pi}(x) = e^{cx}/Z$ for some normalization Z . Hence the symmetry of the rates combined with detailed balance completely constrains the form of the steady state.

Similarly, if the rates $\langle x|\mathbf{W}|y\rangle$ only depend on the ratio x/y (scale invariance), a similar argument shows that $\psi(x) = x^\nu$ for some constant ν . Then the steady state must have the form $\tilde{\pi}(x) = x^\nu/Z$.

1.4.3 Detailed Balance in Physical Systems

It can be shown that detailed balance is required of certain physical systems. Suppose we have a closed, classical system with a Markovian macroscopic stochastic process $Y(\{q\}, \{p\})$ that depends on the microscopic phase space coordinates $\{q\}, \{p\}$. We will show this process obeys detailed balance provided the microscopic Hamiltonian and macroscopic variable Y are both even functions of all momenta.⁵

Denote phase space coordinates of the entire system by $x = (\mathbf{q}_1, \mathbf{p}_1, \dots, \mathbf{q}_N, \mathbf{p}_N)$. Hence the stochastic process can be written $Y(x)$. Define the time-reversal transformation by

$$x \rightarrow \bar{x} = (\mathbf{q}_1, -\mathbf{p}_1, \dots, \mathbf{q}_N, -\mathbf{p}_N) \quad (1.64)$$

[to be continued.....]

⁵If there is a magnetic field or an overall rotation of the system, this argument can be modified to yield a symmetry similar to detailed balance. See van Kampen.

1.5 Entropy and the Approach to Steady State

The existence of a steady state solution to the master equation is obvious, either from physical intuition or from mathematical proof using the theorems of Perron and Frobenius. In this section we investigate how the system approaches the steady state.

Our first task is to construct an entropy function (known more generally as a Lyapunov function). The goal is to show that this function increases monotonically over time. Assume a normalized, nonnegative steady state solution $\tilde{\pi}(x)$. (We assume an irreducible process on a finite state space, which means there are no transient states.) Now take any function $f(k)$ such that $f(k) \geq 0$ and $f''(k) > 0$ for all $k \geq 0$. Hence $f(k)$ is a nonnegative convex function. Now we define the time-dependent function $H(t)$ as

$$H(t) = - \sum_x \tilde{\pi}(x) f(k_{x,t}), \quad \text{where} \quad k_{x,t} = - \frac{\pi(x,t)}{\tilde{\pi}(x)}. \quad (1.65)$$

Note that $H(t) \leq 0$ for all t . We will show that $H(t)$ increases monotonically but can never be positive, and thus it approaches a limit.

So we consider the time derivative:

$$\begin{aligned} \partial_t H(t) &= - \sum_x \tilde{\pi}(x) f'(k_{x,t}) \partial_t k_{x,t} = - \sum_x f'(k_{x,t}) \partial_t \pi(x,t) \\ &= - \sum_x f'(k_{x,t}) \sum_y (\langle x|\mathbf{W}|y\rangle \pi(y,t) - \langle y|\mathbf{W}|x\rangle \pi(x,t)) \\ &= - \sum_{x,y} f'(k_{x,t}) \langle x|\mathbf{W}|y\rangle \pi(y,t) + \sum_{x,y} f'(k_{x,t}) \langle y|\mathbf{W}|x\rangle \pi(x,t) \\ &= - \sum_{x,y} f'(k_{x,t}) \langle x|\mathbf{W}|y\rangle \pi(y,t) + \sum_{x,y} f'(k_{y,t}) \langle x|\mathbf{W}|y\rangle \pi(y,t) \\ &= - \sum_{x,y} (f'(k_{x,t}) - f'(k_{y,t})) \langle x|\mathbf{W}|y\rangle \pi(y,t) \\ &= - \sum_{x,y} (k_{y,t} f'(k_{x,t}) - k_{y,t} f'(k_{y,t})) \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) \end{aligned} \quad (1.66)$$

Now we are going to add some terms that sum to zero. It can be shown that

$$\sum_{x,y} a_{x,t} \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) = \sum_x a_{x,t} \left(\sum_y \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) \right) = 0 \quad (1.67)$$

$$\sum_{x,y} a_{y,t} \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) = \sum_y a_{y,t} \tilde{\pi}(y) \left(\sum_x \langle x|\mathbf{W}|y\rangle \right) = 0 \quad (1.68)$$

Choosing $a_{x,t} = f(k_{x,t}) - k_{x,t} f'(k_{x,t})$, we add the zero sum

$$\begin{aligned} & - \sum_{x,y} \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) (a_{x,t} - a_{y,t}) \\ &= - \sum_{x,y} \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y) ([f(k_{x,t}) - f(k_{y,t})] - [k_{x,t} f'(k_{x,t}) - k_{y,t} f'(k_{y,t})]) \end{aligned} \quad (1.69)$$

to the above expression for dH/dt to get

$$\partial_t H(t) = - \sum_{x,y} [f(k_{x,t}) - f(k_{y,t}) + (k_{y,t} - k_{x,t}) f'(k_{x,t})] \langle x|\mathbf{W}|y\rangle \tilde{\pi}(y). \quad (1.70)$$

By a geometric argument, one can show that the quantity in [...] is always nonpositive, and therefore dH/dt is always nonnegative. So $H(t)$ increases monotonically.

But since $H(t)$ is bounded from above by zero, then it must tend to a limit as $t \rightarrow \infty$, in which case $dH/dt \rightarrow 0$. For pairs of states x, y such that $\langle x | \mathbf{W} | y \rangle \neq 0$, this requires that $k_{x,t} = k_{y,t}$. Since we are assuming irreducibility (one communication class), this means $k_{x,t} = \text{constant}$ for all states. This will only be true if $\pi(x, t)$ is proportional to $\tilde{\pi}$, and the normalization condition then means they are equal.

This argument is for any nonnegative convex function $f(k)$, but in physics we choose $f(k) = k \log k$ yielding the standard entropy functional

$$H(t) = - \sum_x \pi(x, t) \log \left(\frac{\pi(x, t)}{\tilde{\pi}(x)} \right). \quad (1.71)$$

A key reason this function is used is because it makes entropy “extensive” in the sense that entropies of separate systems add. That is, suppose we have system 1 with states $x \in \mathcal{S}_1$ and system 2 with states $y \in \mathcal{S}_2$. Then the entropy of the combined system is

$$\begin{aligned} H_{1+2}(t) &= - \sum_{x \in \mathcal{S}_1} \sum_{y \in \mathcal{S}_2} \pi_1(x, t) \pi_2(y, t) \log \left(\frac{\pi_1(x, t) \pi_2(y, t)}{\tilde{\pi}_1(x) \tilde{\pi}_2(y)} \right) \\ &= - \sum_{x \in \mathcal{S}_1} \pi_1(x, t) \log \left(\frac{\pi_1(x, t)}{\tilde{\pi}_1(x)} \right) - \sum_{y \in \mathcal{S}_2} \pi_2(y, t) \log \left(\frac{\pi_2(y, t)}{\tilde{\pi}_2(y)} \right) \\ &= H_1(t) + H_2(t). \end{aligned} \quad (1.72)$$

Note that this entropy is time-dependent quantity and it actually tends to zero as the system approaches steady state. So strictly speaking, for thermodynamic systems the total “generalized entropy” is of the form $S(t) = kH(t) + S_{eq}$, where k is Boltzmann’s constant and S_{eq} is the constant equilibrium entropy. So $H(t)$ really describes the difference between non-equilibrium and equilibrium entropy.

One particular question of interest is whether the matrix elements of $e^{t\mathbf{W}}$ converge to their steady state values monotonically. This is certainly not true in general – none of the matrix elements from the cyclic 3×3 system converge monotonically. But what if the system obeys detailed balance? Denoting the steady state as

$$|\tilde{\pi}\rangle \equiv \sum_j \tilde{\pi}_j |j\rangle, \quad (1.73)$$

we define detailed balance to be

$$\langle i | \mathbf{W} | j \rangle \tilde{\pi}_j = \langle j | \mathbf{W} | i \rangle \tilde{\pi}_i. \quad (1.74)$$

Now if $a = b$, then the coefficients are strictly positive, and this function must be monotonically decreasing in time. This can be seen by taking the time derivative, and noting that all eigenvalues λ_α are non-positive, so the function must be decreasing at all times.

However, this monotonicity is not guaranteed for the off-diagonal elements, even with detailed balance. Indeed, consider the system

$$\mathbf{W} = \begin{bmatrix} -\mu & \mu & 0 \\ \mu & -\mu - \epsilon & \epsilon \\ 0 & \epsilon & -\epsilon \end{bmatrix}, \quad (1.75)$$

where $\mu > \epsilon$; for instance, take $\mu = 1$ and $\epsilon = 10^{-1}$. Then the off-diagonal terms $\langle 2 | e^{t\mathbf{W}} | 1 \rangle = \langle 1 | e^{t\mathbf{W}} | 2 \rangle$ are non-monotonic. This is because the different relaxation times mean a system starting in state 1 or 2 quickly equilibrates between these two states (since their transitions rates are high), ignoring state 3. But after a longer time they equilibrate with state 3 and the probability drops to its final steady state. Therefore the

existence of monotonic (or nearly monotonic) off-diagonal probabilities depends more subtly on the time scales of the system. In this case, the probability from a “fast” state (1 or 2) to a “slow” state (3) is monotonic, but the probability from a fast state to another fast state is not.

1.6 The Backward Equation and Observables

The master equation describes how the probability distribution of states propagates forward in time. A related equation, the **backward equation**, is obtained by taking the transpose of the rate matrix:

$$\partial_t \langle \phi(t) \rangle = \mathbf{W}^\dagger | \phi(t) \rangle, \quad \partial_t \phi(x, t) = \sum_y \langle x | \mathbf{W}^\dagger | y \rangle \phi(y, t). \quad (1.76)$$

The general solution to this equation is

$$\phi(x, t) = \sum_y \langle x | e^{t\mathbf{W}^\dagger} | y \rangle \phi(y, 0) = \sum_y \phi(y, 0) \langle y | e^{t\mathbf{W}} | x \rangle. \quad (1.77)$$

It is clear that a constant function, $\phi(x, t) = C$, is a steady state solution:

$$\partial_t \phi(x, t) = \sum_y \langle x | \mathbf{W}^\dagger | y \rangle \phi(y, t) = C \sum_y \langle y | \mathbf{W} | x \rangle = 0. \quad (1.78)$$

Alternatively, this can be seen by writing the backward equation as

$$\begin{aligned} \partial_t \phi(x, t) &= \sum_y \langle x | \mathbf{W}^\dagger | y \rangle \phi(y, t) \\ &= \sum_y \langle x | \mathbf{W}^\dagger | y \rangle \phi(y, t) - \phi(x, t) \sum_y \langle y | \mathbf{W} | x \rangle \\ &= \sum_y \langle x | \mathbf{W}^\dagger | y \rangle (\phi(y, t) - \phi(x, t)), \end{aligned} \quad (1.79)$$

where we have added $0 = \sum_y \langle y | \mathbf{W} | x \rangle$ in the second line. Now this makes it clear that for a finite, discrete state space, the largest component of ϕ must decrease and the smallest component must increase. This means all components must tend to a single constant as $t \rightarrow \infty$.

One chief use of the backward equation is to describe observables. Let \mathcal{O} be a function that maps each state x to some observable quantity $\mathcal{O}(x)$. Assume it is numerical so we may define the average value of \mathcal{O} at time t :

$$\langle \mathcal{O} \rangle_t = \sum_x \mathcal{O}(x) \pi(x, t). \quad (1.80)$$

Inserting the formal solution of the master equation for $\pi(x, t)$,

$$\begin{aligned} \langle \mathcal{O} \rangle_t &= \sum_x \mathcal{O}(x) \pi(x, t) \\ &= \sum_x \mathcal{O}(x) \sum_y \langle x | e^{t\mathbf{W}} | y \rangle \pi(y, 0) \\ &= \sum_y \left(\sum_x \mathcal{O}(x) \langle x | e^{t\mathbf{W}} | y \rangle \right) \pi(y, 0) \\ &= \sum_y \mathcal{O}(y, t) \pi(y, 0) \\ &= \langle \mathcal{O}(t) \rangle_0. \end{aligned} \quad (1.81)$$

It is clear that $\mathcal{O}(y, t) \equiv \sum_x \mathcal{O}(x) \langle x | e^{t\mathbf{W}} | y \rangle$ is a solution to the backward equation. This procedure is equivalent to the transition between the Schrödinger picture (observables are time-independent while state vectors evolve in time) and the Heisenberg picture (state vectors are time-independent while observables evolve in time) of quantum mechanics. Apparently observables are solutions of the backward equation, and their infinite time limit is a constant!

This yields an interesting angle on proving the convergence to a steady state distribution for $\pi(x, t)$. Suppose we have a “Heisenberg picture” observable $\mathcal{O}(x, t)$ that tends to the constant C as $t \rightarrow \infty$. Hence

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_x \mathcal{O}(x, t) \pi(x, 0) &= C \sum_x \pi(x, 0) = C \\ &= \lim_{t \rightarrow \infty} \sum_x \mathcal{O}(x) \pi(x, t). \end{aligned} \tag{1.82}$$

[How does this argue for a steady state limit of $\phi(x, t)$?

1.7 Martingales

Martingales are a special class of stochastic processes with the property

$$\mathbb{E}[X(t+1) | X(t) = x] = x. \tag{1.83}$$

That is, if the system is in state x now, then its expected value in the next step is also x . Thus, it is a sort of “unbiased” process. Indeed, this also means that the Fokker-Planck equation for the process has no drift term (no first derivative term), since that term contains the first jump moment, which is zero for a martingale. Moreover, this doesn’t only related the state now and the state one step from now. Since for a discrete time process,

$$\mathbb{E}[X(t+1) | X(t) = x] = \sum_x x \langle x | \mathbf{P} | y \rangle = y, \tag{1.84}$$

we can multiply both sides by $\langle y | \mathbf{P} | z \rangle$ and sum over y to obtain

$$\begin{aligned} \sum_y \sum_x x \langle x | \mathbf{P} | y \rangle \langle y | \mathbf{P} | z \rangle &= \sum_y y \langle y | \mathbf{P} | z \rangle \\ \sum_x x \langle x | \mathbf{P}^2 | z \rangle &= z = \mathbb{E}[X(t+2) | X(t) = z]. \end{aligned} \tag{1.85}$$

Hence it is generally true (for both discrete and continuous time processes)

$$\mathbb{E}[X(t + \Delta t) | X(t) = x] = x. \tag{1.86}$$

So if the system is in state x at some point, then the expected state at *any* later time is also x .

One property of martingales is that they must have multiple steady states. This is because the martingale property is

$$\sum_x x \langle x | \mathbf{P} | y \rangle = y, \tag{1.87}$$

which means there is a left eigenvector $\sum_x x \langle x |$ with eigenvalue 1. But we also know there is another left eigenvector with eigenvalue 1, $\sum_x \langle x |$. Therefore there must also be two linearly independent right eigenvectors with eigenvalue 1, and hence two linearly independent steady states. [See Wright-Fisher or Moran model for examples.]

This can be used to argue that the fixation of a neutral mutation in any martingale population model must be x_0/N , the initial mutant frequency. Suppose we have a martingale on a state space $X = 0, 1, \dots, N$, and 0 and N are absorbing states. Then

$$x_0 = \sum_x x \langle x | \mathbf{\Pi}(\infty) | x_0 \rangle = N \langle N | \mathbf{\Pi}(\infty) | x_0 \rangle \quad (1.88)$$

and hence $\langle N | \mathbf{\Pi}(\infty) | x_0 \rangle = x_0/N$.

1.8 The Deterministic Limit

The master equation describes how the full probability distribution of some stochastic process evolves over time. However, it is often useful to consider the limit when the stochastic fluctuations are small and the stochastic quantity becomes deterministic in its evolution.

Suppose our stochastic process $X(t)$ takes on numerical values in some state space \mathcal{S} . Then we can define averages

$$\langle X \rangle_t = \sum_x x \pi(x, t). \quad (1.89)$$

(Here X is used as a Schrödinger picture observable.) Now this quantity satisfies the differential equation:

$$\begin{aligned} \partial_t \langle X \rangle_t &= \sum_x x \partial_t \pi(x, t) \\ &= \sum_x x \sum_y (\langle x | \mathbf{W} | y \rangle \pi(y, t) - \langle y | \mathbf{W} | x \rangle \pi(x, t)) \\ &= \sum_{x,y} (x - y) \langle x | \mathbf{W} | y \rangle \pi(y, t). \end{aligned} \quad (1.90)$$

Define the n th **jump moment** $m_n(y)$ as the n th moment of the probability distribution of jumps starting from y :

$$m_n(y) = \sum_x (x - y)^n \langle x | \mathbf{W} | y \rangle. \quad (1.91)$$

Hence $m_1(y)$ describes the mean change in X in one step starting from $X = y$. Therefore the deterministic equation becomes

$$\partial_t \langle X \rangle_t = \sum_y m_1(y) \pi(y, t) = \langle m_1(X) \rangle_t, \quad (1.92)$$

where $m_1(X)$ is a Schrödinger picture observable.

Assuming $m_1(X)$ is analytic function of X , we can expand in powers:

$$\partial_t \langle X \rangle_t = \langle m_1(X) \rangle_t = m_1(\langle X \rangle_t) + \frac{1}{2} \langle (X - \langle X \rangle_t)^2 \rangle_t m_1''(\langle X \rangle_t) + \dots \quad (1.93)$$

This is now a differential equation for the deterministic time evolution of $\langle X \rangle_t$. Often higher order terms in the expansion can be ignored under certain limits and the deterministic equation reduces to

$$\partial_t \langle X \rangle_t = m_1(\langle X \rangle_t). \quad (1.94)$$

This is often justified by taking the system size to infinity, which usually causes the higher order fluctuation terms to disappear.

2 Markov Processes on 1D Lattices

An important class of Markov processes occur on 1D lattices. These are often called birth-death or generation-recombination processes, although N. G. van Kampen’s less loaded term of “one-step” processes seems more appropriate in general. The important aspect of these processes is that jumps can only occur between nearest neighbors on the lattice, a property which both arises very naturally in many applications and also makes them much simpler to study. They can be continuous or discrete in time, but here we will focus on continuous time. Most results are essentially the same in both cases. *[Insert figure with 1D lattice and rates.]*

Since the states are a 1D lattice with only nearest-neighbor connectivity, the state space can be considered on \mathbb{Z} , with the states indexed by an integer $n \in \mathbb{Z}$. Hence the only nonzero elements of the rate matrix are

$$\begin{aligned} \langle n+1 | \mathbf{W} | n \rangle &= u_n \\ \langle n-1 | \mathbf{W} | n \rangle &= d_n \\ \langle n | \mathbf{W} | n \rangle &= -u_n - d_n, \end{aligned} \tag{2.1}$$

where u_n is the rate “up” from n and d_n is the rate “down” from n . Hence \mathbf{W} is a tridiagonal matrix – as a result some mathematicians call such processes “continuants” in reference to the polynomials arising in these matrices’ determinants. State spaces on \mathbb{Z} will be in one of three classes:

- Class I: finite lattice. Without loss of generality, we can let $n \in \{0, 1, \dots, N\}$.
- Class II: half-infinite lattice ($n \in \{0, 1, \dots, \infty\}$).
- Class III: infinite lattice ($n \in \{-\infty, \dots, -1, 0, 1, \dots, \infty\}$).

In all cases the master equation will be of the form

$$\partial_t \pi_n(t) = u_{n-1} \pi_{n-1}(t) + d_{n+1} \pi_{n+1}(t) - (u_n + d_n) \pi_n(t). \tag{2.2}$$

For classes I and II, some modification of this equation may be necessary for n near or on the boundaries.

2.1 General Properties

2.1.1 Moments and Deterministic Equation

The moment functions as defined in Equation 1.91 are

$$m_j(n) = u_n + (-1)^j d_n. \tag{2.3}$$

Hence the deterministic equation is

$$\partial_t \mathbb{E}[\mathbf{N}(t)] = \mathbb{E}[m_1(\mathbf{N})] = \mathbb{E}[u_{\mathbf{N}}] - \mathbb{E}[d_{\mathbf{N}}]. \tag{2.4}$$

2.1.2 Creation and Annihilator Operators

Since these processes move only one step at a time up and down a lattice, it is sometimes convenient to study them in terms of creation and annihilation operators. Define function operators⁶

$$\mathbf{a}^\dagger f(n) = f(n+1), \quad \mathbf{a} f(n) = f(n-1). \tag{2.5}$$

Some redefinition is required when the range of n is not infinite. In any case, we can now rewrite the master equation as

⁶These aren’t quite the same kind of objects as used in quantum mechanics – they are normalized differently and have commutator zero instead of one. For the kinds of processes in this section, this is a convenient definition, but other definitions (some more analogous to that of quantum mechanics) can be used in other settings.

$$\partial_t \pi_n(t) = [(\mathbf{a}^\dagger - 1)d_n + (\mathbf{a} - 1)u_n]\pi_n(t). \quad (2.6)$$

The creation/annihilation operators act on u_n and d_n as functions of n .

2.1.3 Currents and Steady State

Creation and annihilation operators can be used to deduce the steady state distribution very easily. The steady state distribution $\tilde{\pi}_n$ satisfies

$$\begin{aligned} 0 &= [(\mathbf{a}^\dagger - 1)d_n + (\mathbf{a} - 1)u_n]\tilde{\pi}_n \\ &= (\mathbf{a}^\dagger - 1)(d_n - \mathbf{a}u_n)\tilde{\pi}_n. \end{aligned} \quad (2.7)$$

This implies that $(d_n - \mathbf{a}u_n)\tilde{\pi}_n$ is some n -independent quantity $-J$, so-called because it represents the net probability current from n to $n - 1$:

$$J = u_{n-1}\tilde{\pi}_{n-1} - d_n\tilde{\pi}_n = \mathbf{a}u_n\tilde{\pi}_n - d_n\tilde{\pi}_n. \quad (2.8)$$

For state space classes I and II, when there is a boundary at $n = 0$, the steady-state master equation for $n = 1$ implies that

$$u_0\tilde{\pi}_0 - d_1\tilde{\pi}_1 = 0. \quad (2.9)$$

Hence $J = 0$ everywhere; this is not surprising since for probability to be conserved, net current has nowhere to go at a boundary. Therefore

$$u_{n-1}\tilde{\pi}_{n-1} = d_n\tilde{\pi}_n \quad (2.10)$$

for all n , i.e., detailed balance or time reversibility is satisfied. This allows one to deduce the steady state:

$$\tilde{\pi}_n = \frac{u_{n-1}u_{n-2}\cdots u_0}{d_n d_{n-1}\cdots d_1} \tilde{\pi}_0, \quad (2.11)$$

where $\tilde{\pi}_0$ is chosen using normalization. If there is no boundary (class III), then we take as an ansatz the above for $n > 0$ and the following for $n < 0$:

$$\tilde{\pi}(n) = \frac{d_{n+1}d_{n+2}\cdots d_0}{u_n u_{n+1}\cdots u_{-1}} \tilde{\pi}(0). \quad (2.12)$$

Substitution of this ansatz into the master equation shows that it is indeed a steady state for all $n \in \mathbb{Z}$. However, in the case of an infinite range, J may not be zero and hence the model may not be reversible: there is a constant current from $-\infty$ to $+\infty$ which distinguishes the flow of time. In this case, a separate but similar steady state equation may be derived.

2.2 Linear Rates

In general, the rates u_n and d_n can be any numbers. However, in many cases they will depend linearly on n :

$$u_n = \alpha n + u_0, \quad d_n = \beta n + d_0. \quad (2.13)$$

Such cases arise very naturally in many systems, particularly those that describe noninteracting degrees of freedom (more on this later). Their simplicity also allows for a general solution to the master equation, which we derive here.

2.2.1 Exact Solution

We first let n take all values in \mathbb{Z} to facilitate calculations (we will address this afterward). Assume $\alpha, \beta \neq 0$ and $\alpha \neq \beta$. Then the master equation is

$$\partial_t \pi_n(t) = (\beta(n+1) + d_0)\pi_{n+1}(t) + (\alpha(n-1) + u_0)\pi_{n-1}(t) - ((\alpha + \beta)n + u_0 + d_0)\pi_n(t). \quad (2.14)$$

A convenient tool for solving many one-step master equations is the probability generating function, defined as

$$F(z, t) = \sum_n z^n \pi_n(t). \quad (2.15)$$

It is the same as the characteristic function with $z = e^{ik}$. Its derivatives generate moments, and we can obtain the full distribution by either identifying coefficients in its power series or by an explicit inversion:

$$\pi(n, t) = \frac{1}{2\pi i} \oint dz z^{-n-1} F(z, t), \quad (2.16)$$

where the contour is the unit circle.

We can convert the master equation to an equation for the generating function $F(z, t)$ by multiplying both sides by z^n and summing over n . For Equation 2.14 this gives

$$\partial_t F(z, t) = (\alpha z - \beta)(z - 1)\partial_z F(z, t) + (z - 1) \left(u_0 - \frac{d_0}{z} \right) F(z, t). \quad (2.17)$$

We can solve this using the method of characteristics. Rewriting in the form

$$\partial_t F(z, t) - (\alpha z - \beta)(z - 1)\partial_z F(z, t) = (z - 1) \left(u_0 - \frac{d_0}{z} \right) F(z, t), \quad (2.18)$$

we see that the Lagrange-Charpit equations are

$$dt = -\frac{dz}{(\alpha z - \beta)(z - 1)} = \frac{dF}{(z - 1)(u_0 - d_0/z)F}. \quad (2.19)$$

The first can be integrated to give the constant C_1 ,

$$C_1 = \left(\frac{z - 1}{\alpha z - \beta} \right) e^{(\alpha - \beta)t}, \quad (2.20)$$

which must be constant along the characteristic curves. The second can also be integrated to give

$$F(z, t) = C_2 z^{-d_0/\beta} (\beta - \alpha z)^{d_0/\beta - u_0/\alpha}. \quad (2.21)$$

The constant C_2 can be an arbitrary function of the constant C_1 :

$$F(z, t) = z^{-d_0/\beta} (\beta - \alpha z)^{d_0/\beta - u_0/\alpha} \Omega \left(\left(\frac{z - 1}{\alpha z - \beta} \right) e^{(\alpha - \beta)t} \right). \quad (2.22)$$

We can determine this function by imposing initial conditions. Suppose that the initial state is n_0 . Then the initial condition for the generating function is $F(z, 0) = z^{n_0}$. This shows

$$\Omega \left(\frac{z - 1}{\alpha z - \beta} \right) = z^{n_0 + d_0/\beta} (\beta - \alpha z)^{u_0/\alpha - d_0/\beta}. \quad (2.23)$$

We can define a new variable ζ as

$$\zeta = \frac{z-1}{\alpha z - \beta}, \quad \rightarrow \quad z = \frac{\beta\zeta - 1}{\alpha\zeta - 1}, \quad (2.24)$$

and then write

$$\Omega(\zeta) = \left(\frac{\beta\zeta - 1}{\alpha\zeta - 1} \right)^{n_0 + d_0/\beta} \left(\frac{\alpha - \beta}{\alpha\zeta - 1} \right)^{u_0/\alpha - d_0/\beta}. \quad (2.25)$$

This yields the full solution for $F(z, t)$:

$$F(z, t) = z^{n_0} (\beta - \alpha)^{u_0/\alpha - d_0/\beta} \times \left(\frac{\beta}{z} (1 - e^{(\alpha-\beta)t}) - \alpha + \beta e^{(\alpha-\beta)t} \right)^{n_0 + d_0/\beta} (-\alpha z (1 - e^{(\alpha-\beta)t}) + \beta - \alpha e^{(\alpha-\beta)t})^{-u_0/\alpha - n_0}. \quad (2.26)$$

Now we can impose some more realistic conditions on this. In particular, we must ensure that the rates are always nonnegative and the outward rates at any boundaries are zero. We can consider this in each of the three state space classes:

- Class I: in this case $d_n = \beta n$ and $u_n = \alpha(N - n)$. The general solution here is

$$F(z, t) = \frac{1}{(\alpha + \beta)^N} (\beta(1 - e^{-(\alpha+\beta)t}) + z(\alpha + \beta e^{-(\alpha+\beta)t}))^{n_0} (\alpha z(1 - e^{-(\alpha+\beta)t}) + \beta + \alpha e^{-(\alpha+\beta)t})^{N - n_0}. \quad (2.27)$$

- Class II: $d_n = \beta n$ and $u_n = \alpha n + u_0$. The general solution in this case is

$$F(z, t) = (\beta - \alpha)^{u_0/\alpha} \times \left(\beta(1 - e^{(\alpha-\beta)t}) + z(\beta e^{(\alpha-\beta)t} - \alpha) \right)^{n_0} (\alpha z(e^{(\alpha-\beta)t} - 1) + \beta - \alpha e^{(\alpha-\beta)t})^{-u_0/\alpha - n_0}. \quad (2.28)$$

- Class III: the rate coefficients must be constant ($u_n = \alpha$, $d_n = \beta$). This is just a random walk, which will be solved in the examples.

These generating functions can be inverse transformed to obtain the probability distributions, although in general they are complicated. In the next section we discuss a physical interpretation of the class I linear rate models which allows a more transparent derivation of $\pi_n(t)$.

2.2.2 Collective Systems

As an aside and partial motivation to considering one-step processes with linear rates, let us consider the following class of models. Suppose we have a single degree of freedom that evolves according to the master equation

$$\partial_t \pi_1(m', t) = \sum_m \langle m' | \mathbf{W}_1 | m \rangle \pi_1(m, t), \quad (2.29)$$

where the subscript 1 indicates that these quantities are for this single DOF. Suppose we have a collection of N such systems. The total probability distribution of this collection is given by the product of single distributions:

$$\prod_i \pi_1(m_i, t). \quad (2.30)$$

But suppose we don't care about the precise specification of the whole collection $\{m_1, \dots, m_N\}$, but we only care about the occupation numbers – that is, the number of these systems in the m th state: $n_m \in \{0, \dots, N\}$. Denote the set of occupation numbers by $\{n\}$, which contains as many occupation numbers as there are states for the individual degrees of freedom (possibly infinite). We can obtain the master equation for these occupation numbers as follows. For a sufficiently small time interval Δt , the only possible change in the occupation numbers is due to a single system moving from a state m to state m' . Thus the master equation is

$$\begin{aligned}\partial_t P(\{n\}, t) &= \sum_{m, m'} \langle m' | \mathbf{W}_1 | m \rangle [(n_m + 1)P(\{\dots, n_m + 1, n_{m'} - 1, \dots\}, t) - n_m P(\{n\}, t)] \\ &= \sum_{m, m'} \langle m' | \mathbf{W}_1 | m \rangle (\mathbf{a}_{m'} \mathbf{a}_m^\dagger - 1) n_m P(\{n\}, t).\end{aligned}\tag{2.31}$$

The physics of the occupation number master equation is obviously the same as that of the individual system master equations, so we can express the solution of the former in terms of the latter. Suppose that all the systems begin in the same level m , so that the initial occupation numbers are $n_m = N$, $n_{m'} = 0$ for $m' \neq m$. Then the occupation number solution should be

$$P(\{n\}, t) = \binom{N}{n_1, n_2, \dots} \prod_j \langle j | e^{t \mathbf{W}_1} | m \rangle^{n_j}.\tag{2.32}$$

As we would want to do in many cases, if the total number of systems N is distributed in some way (such as in the grand canonical ensemble), we can sum over these expressions for N and eliminate the constraints on the occupation numbers.

As an example, consider a collection of two-state systems, where each is “on” (1) or “off” (0). Their rate matrix elements are

$$\langle 1 | \mathbf{W}_1 | 0 \rangle = \alpha, \quad \langle 0 | \mathbf{W}_1 | 1 \rangle = \beta.\tag{2.33}$$

There are only two occupation numbers, n_0 and n_1 . The full master equation is

$$\partial_t P(n_0, n_1, t) = [\langle 0 | \mathbf{W}_1 | 1 \rangle (\mathbf{a}_0 \mathbf{a}_1^\dagger - 1) n_1 + \langle 1 | \mathbf{W}_1 | 0 \rangle (\mathbf{a}_1 \mathbf{a}_0^\dagger - 1) n_0] P(n_0, n_1, t).\tag{2.34}$$

However, since $n_0 = N - n_1$, we can rewrite this as (with $n = n_1$ the number of on systems)

$$\partial_t P(n, t) = [\alpha (\mathbf{a} - 1) (N - n) + \beta (\mathbf{a}_1^\dagger - 1) n] P(n, t).\tag{2.35}$$

Thus we see how linear rates in a one-step process arise very naturally from considering the master equation for the occupation number for a collection of noninteracting two-state systems. Nonlinear dependencies on the state variable n in the rates are indicative of interactions, which obviously makes them more difficult to solve.

We use this picture to solve the linear rates problem for a class I state space (rather than by the generating function obtained in the previous section). The single DOF time-evolution operator is

$$e^{t \mathbf{W}_1} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha e^{-(\alpha + \beta)t} & \beta(1 - e^{-(\alpha + \beta)t}) \\ \alpha(1 - e^{-(\alpha + \beta)t}) & \alpha + \beta e^{-(\alpha + \beta)t} \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.\tag{2.36}$$

We see that the generating function of the previous section is just

$$F(z, t) = (p_{01} + z p_{11})^{n_0} (p_{00} + z p_{10})^{N - n_0}.\tag{2.37}$$

Assuming that we initially have n_0 of the DOFs in the “on” state, the probability of having n “on” at time t is

$$\pi_n(t) = \sum_{i=0}^{n_0} \sum_{j=0}^{N-n_0} \delta_{i+j,n} \left[\binom{n_0}{i} p_{11}^i p_{01}^{n_0-i} \right] \left[\binom{N-n_0}{j} p_{10}^j p_{00}^{N-n_0-j} \right]. \quad (2.38)$$

Here we see that the index i refers to how many of the initially “on” states are “on” again at time t (meaning $n_0 - i$ are now “off”) and j refers to how many of the initially “off” states are now “on” at time t (meaning $N - n_0 - j$ are again “off”). This obviously is the same expression that would be obtained from expanding $F(z, t)$ in powers of z and picking out the n th coefficient.

The steady state of the single DOF time-evolution operator is

$$\frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \beta \\ \alpha & \alpha \end{pmatrix}. \quad (2.39)$$

Defining $\gamma = \alpha/\beta$,

$$\begin{aligned} \tilde{\pi}_n &= \frac{1}{(1 + \gamma)^N} \sum_{i=0}^{n_0} \sum_{j=0}^{N-n_0} \delta_{i+j,n} \binom{n_0}{i} \binom{N-n_0}{j} \gamma^{i+j} \\ &= \frac{\gamma^n}{(1 + \gamma)^N} \sum_{i=0}^{n_0} \sum_{j=0}^{N-n_0} \delta_{i+j,n} \binom{n_0}{i} \binom{N-n_0}{j} \\ &= \frac{1}{(1 + \gamma)^N} \binom{N}{n} \gamma^n. \end{aligned} \quad (2.40)$$

This agrees with Equation 2.11. Thus, in general class I linear rate models always have binomial steady-state distributions.

2.3 Spectral Expansion

Spectral expansions of one-step processes are also possible, albeit sometimes technically difficult. The existence of time reversibility (with the constant current case exception) guarantees a spectrum of real, nonpositive eigenvalues of \mathbf{W} and a complete basis of eigenvectors, as shown earlier.

Cases where explicit spectral expansions are possible include random walks on finite and half-infinite lattices, where the generating function method encounters difficulties due to the boundary conditions. First we consider the half-infinite (class II) case – an asymmetric random walk with boundary at $n = 0$:

$$\begin{aligned} \partial_t \pi_n(t) &= \alpha \pi_{n-1}(t) + \beta \pi_{n+1}(t) - (\alpha + \beta) \pi_n(t), \quad (n \geq 1) \\ \partial_t \pi_0(t) &= \beta \pi_1(t) - \alpha \pi_0(t). \end{aligned} \quad (2.41)$$

Let the spectrum of eigenvalues be $-\lambda$, where $\lambda \geq 0$. Then the components of the eigenvectors for λ will be denoted by $\psi_n^{(\lambda)}$. Therefore these will satisfy

$$\begin{aligned} (\alpha + \beta - \lambda) \psi_n^{(\lambda)} &= \alpha \psi_{n-1}^{(\lambda)} + \beta \psi_{n+1}^{(\lambda)} \\ (\alpha - \lambda) \psi_0^{(\lambda)} &= \beta \psi_1^{(\lambda)}. \end{aligned} \quad (2.42)$$

We take the standard ansatz $\psi_n^{(\lambda)} = z_\lambda^n$, where z_λ must satisfy

$$\alpha + \beta - \lambda = \beta z_\lambda + \frac{\alpha}{z_\lambda}, \quad \longrightarrow \quad z_\lambda = \frac{\alpha + \beta - \lambda \pm \sqrt{(\lambda - \alpha - \beta)^2 - 4\alpha\beta}}{2\beta}. \quad (2.43)$$

Note that the product of the two solutions, z_λ and z'_λ , is α/β . The general solution for the eigenvectors is then a linear combination of these solutions (ignoring the irrelevant overall normalization):

$$\psi_n^{(\lambda)} = z_\lambda^n + C_\lambda z_\lambda'^n, \quad (2.44)$$

where C_λ is constant in n . This is the form of $\psi_n^{(\lambda)}$ for $n \geq 1$, and we can determine the value of C_λ by using the equations for $\psi_0^{(\lambda)}$:

$$\begin{aligned} \psi_0^{(\lambda)} &= \frac{\beta}{\alpha - \lambda} (z + C_\lambda z') \\ C_\lambda &= -\frac{\alpha - \beta - \lambda - \sqrt{(\lambda - \alpha - \beta)^2 - 4\alpha\beta}}{\alpha - \beta - \lambda + \sqrt{(\lambda - \alpha - \beta)^2 - 4\alpha\beta}} = -\frac{z' - 1}{z - 1}. \end{aligned} \quad (2.45)$$

Since we don't want the eigenvectors to grow without bound for increasing n , we can let

$$z = \sqrt{\frac{\alpha}{\beta}} e^{i\theta}, \quad z' = \sqrt{\frac{\alpha}{\beta}} e^{-i\theta}, \quad (\pi \geq \theta \geq 0). \quad (2.46)$$

where

$$\theta = \arctan \left(\frac{\sqrt{4\alpha\beta - (\lambda - \alpha - \beta)^2}}{\alpha + \beta - \lambda} \right), \quad \lambda = \alpha + \beta - \sqrt{\alpha\beta} \cos \theta, \quad (2.47)$$

and $2\sqrt{\alpha\beta} + \alpha + \beta > \lambda > 0$. So in principle we now have the complete spectrum of eigenvalues and eigenvectors (now parameterized by $\theta \in [0, \pi]$):

$$\begin{aligned} \psi_0^{(\theta)} &= \frac{\sqrt{\alpha/\beta}}{\sqrt{\alpha/\beta} \cos \theta - 1} (e^{i\theta} + C_\theta e^{-i\theta}) \\ \psi_n^{(\theta)} &= \left(\frac{\alpha}{\beta} \right)^{n/2} (e^{in\theta} + C_\theta e^{-in\theta}), \quad (n \geq 1) \\ \text{where } C_\theta &= -\frac{\sqrt{\alpha/\beta} e^{-i\theta} - 1}{\sqrt{\alpha/\beta} e^{i\theta} - 1}. \end{aligned} \quad (2.48)$$

So the propagator is spectrally expanded as

$$\langle m | e^{t\mathbf{W}} | n \rangle = (1 - \alpha/\beta) \left(\frac{\beta}{\alpha} \right)^n \int_0^\pi d\theta \psi_m^{(\theta)} \psi_n^{(\theta)} e^{-(\alpha + \beta - \sqrt{\alpha\beta} \cos \theta)t}. \quad (2.49)$$

2.4 Examples

Here are several examples of one-step processes with both linear and nonlinear rates.

2.4.1 The Poisson Process

The Poisson process can be considered as a rather trivial example of a one-step process. In this case, the stochastic process $\mathbf{N}(t)$ denotes the number of events that have occurred by time t , if the probability of an event occurring in time Δt is $\alpha \Delta t$. Hence $\mathbf{N}(t)$ takes values $0, 1, 2, \dots$ (class II) and is nondecreasing. The matrix elements are

$$\langle n+1 | \mathbf{W} | n \rangle = u_n = \alpha, \quad \langle n-1 | \mathbf{W} | n \rangle = d_n = 0, \quad (2.50)$$

which gives the master equation

$$\partial_t \pi(n, t) = \alpha(\pi(n-1, t) - \pi(n, t)). \quad (2.51)$$

For the initial condition $\pi(n, 0) = \delta_{n,0}$, it is simple to show that this equation is satisfied by the standard Poisson distribution

$$\pi(n, t) = \frac{(\alpha t)^n}{n!} e^{-\alpha t}. \quad (2.52)$$

In general, the propagator is

$$\langle n' | e^{t\mathbf{W}} | n \rangle = \frac{(\alpha t)^{n'-n}}{(n'-n)!} e^{-\alpha t} \quad (2.53)$$

for $n' \geq n$ and zero otherwise. If α becomes time-dependent, then this generalizes to

$$\pi(n, t) = \frac{1}{n!} \left(\int_0^t d\tau \alpha(\tau) \right)^n e^{-\int_0^t d\tau \alpha(\tau)}. \quad (2.54)$$

2.4.2 Symmetric and Asymmetric Random Walks

For the symmetric random walk on \mathbb{Z} (class III), the rates are (ignoring an overall time scale):

$$\langle n+1 | \mathbf{W} | n \rangle = \langle n-1 | \mathbf{W} | n \rangle = 1. \quad (2.55)$$

Without loss of generality, assume the initial condition is $\pi_n(0) = \delta_{n,0}$. The master equation is

$$\partial_t \pi_n(t) = \pi_{n+1}(t) + \pi_{n-1}(t) - 2\pi_n(t). \quad (2.56)$$

We can solve this using the generating function. Transforming the master equation gives:

$$\partial_t F(z, t) = (z + z^{-1} - 2)F(z, t). \quad (2.57)$$

This can be integrated to give the solution

$$F(z, t) = C(z) e^{t(z+z^{-1}-2)}. \quad (2.58)$$

The arbitrary function $C(z)$ can be fixed by imposing the initial condition $F(z, 0) = 1$, which gives $C(z) = 1$. Expanding in powers of z one can identify the probability distribution

$$\pi_n(t) = e^{-2t} \sum_{\ell} \frac{t^{2\ell+n}}{(\ell+n)! \ell!} = e^{-2t} I_{|n|}(2t). \quad (2.59)$$

where the sum is over all ℓ such that $\ell \geq 0$ and $\ell+n \geq 0$, and I_n is the modified Bessel function.

We can obtain an asymptotic form of this in the limit of $t \rightarrow \infty$ and $n \rightarrow \infty$ with n^2/t fixed. We explicitly invert the generating function using Equation 2.16:

$$\pi_n(t) = \frac{1}{2\pi i} \oint dz z^{-n-1} e^{t(z+z^{-1}-2)} = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-in\theta} e^{2t(\cos\theta-1)}. \quad (2.60)$$

Now we invoke a saddle-point approximation valid in this limit to obtain

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\theta e^{-in\theta - t\theta^2} = \frac{1}{\sqrt{4\pi t}} e^{-\frac{n^2}{4t}}. \quad (2.61)$$

This is of course the expected Gaussian distribution that will also arise from the Fokker-Planck approximation later.

Now we consider the asymmetric random walk:

$$\partial_t \pi_n(t) = \alpha \pi_{n-1}(t) + \beta \pi_{n+1}(t) - (\alpha + \beta) \pi_n(t). \quad (2.62)$$

If we map

$$\pi_n(t) = \phi_n(t) \left(\frac{\alpha}{\beta} \right)^{n/2} e^{-(\alpha + \beta - 2\sqrt{\alpha\beta})t}, \quad (2.63)$$

this gives a new master equation

$$\partial_t \phi_n(t) = \sqrt{\alpha\beta} \phi_{n+1}(t) + \sqrt{\alpha\beta} \phi_{n-1}(t) - 2\sqrt{\alpha\beta} \phi_n(t), \quad (2.64)$$

which is a symmetric random walk that we know how to solve. Therefore the solution is

$$\pi_n(t) = e^{-(\alpha + \beta)t} \left(\frac{\alpha}{\beta} \right)^{n/2} I_{|n|}(2\sqrt{\alpha\beta}t). \quad (2.65)$$

2.4.3 Quantum Harmonic Oscillator and Radiation Field

Consider radiation incident on a quantum harmonic oscillator. The oscillator has states $n = 0, 1, 2, \dots$, with energies $\hbar\omega(n + 1/2)$. The transition rates are

$$\langle n + 1 | \mathbf{W} | n \rangle = u_n = \alpha(n + 1), \quad \langle n - 1 | \mathbf{W} | n \rangle = d_n = \beta n, \quad (2.66)$$

since the rates between n and $n + 1$ are proportional to the matrix element of the dipole moment, which is always proportional to $n + 1$. The steady state is

$$\tilde{\pi}_n = \frac{(\alpha/\beta)^n}{1 - \alpha/\beta}, \quad (2.67)$$

assuming $\alpha < \beta$ (otherwise there is no steady state). From statistical mechanics we know the equilibrium distribution is

$$\tilde{\pi}_n = \frac{1}{Z} e^{-n\hbar\omega/kT}. \quad (2.68)$$

Hence we have

$$e^{-\hbar\omega/kT} = \frac{\alpha}{\beta}. \quad (2.69)$$

2.4.4 Noninteracting Gas

Suppose N total particles of a noninteracting gas occupy two volumes, V_1 and V_2 , connected through a small aperture. After the system equilibrates, what is the probability of there being n particles in V_1 ? Clearly this is

$$\binom{N}{n} \left(\frac{V_1}{V_1 + V_2} \right)^n \left(1 - \frac{V_1}{V_1 + V_2} \right)^{N-n} = \frac{1}{(1 + \gamma)^N} \binom{N}{n} \gamma^n, \quad (2.70)$$

where $\gamma = V_1/V_2$.

Now let us model it as a one-step stochastic process. Let $\mathbf{N}(t)$ be the number of particles in V_1 at time t . Then the rate at which particles are removed should be $\beta\mathbf{N}$ and the rate at which particles are added should be $\alpha(N - \mathbf{N})$. Hence $u_n = \alpha(N - n)$ and $d_n = \beta n$, and the master equation is

$$\partial_t \pi_n(t) = \beta(n + 1)\pi_{n+1}(t) + \alpha(N - (n - 1))\pi_{n-1}(t) - (\beta n + \alpha(N - n))\pi_n(t). \quad (2.71)$$

The steady state distribution is therefore

$$\tilde{\pi}_n = \tilde{\pi}_0 \binom{N}{n} \left(\frac{\alpha}{\beta}\right)^n = \frac{1}{(1 + \alpha/\beta)^N} \binom{N}{n} \left(\frac{\alpha}{\beta}\right)^n. \quad (2.72)$$

Hence $\gamma = \alpha/\beta = V_1/V_2$. As this is a class I linear process, the full solution to the master equation is given by Equation 2.38.

This becomes the grand canonical ensemble in the limit that $N, V_2 \rightarrow \infty$ holding the bath density $\rho = N/V_2$ is fixed. Then the rates become $u_n = \rho V_1$ and $d_n = n$. Hence the steady state distribution is

$$\tilde{\pi}_n = \frac{(\rho V_1)^n}{n!} e^{-\rho V_1}. \quad (2.73)$$

The same model can be used to describe a large bath of molecules at density ρ that can bind to N sites, such as transcription factor proteins and DNA binding sites in the cell nucleus. In this case n is the number of occupied sites, and the rates are $u_n = \alpha(N - n)$ and $d_n = \beta n$. The statistical steady state expression is as before; the equilibrium distribution from statistical mechanics is:

$$\tilde{\pi}_n = \frac{1}{Z} \binom{N}{n} \zeta^n e^{(-n\epsilon + n\mu)/kT}, \quad (2.74)$$

where ζ is the internal partition function of the molecule, ϵ is the energy difference between bound and unbound states, and $\mu = kT \log(\rho \lambda^3)$ is the chemical potential. Thus

$$\frac{\alpha}{\beta} = \zeta \rho \lambda^3 e^{-\epsilon/kT}. \quad (2.75)$$

2.4.5 The Moran Population Process

We now consider a few nonlinear rate processes. Suppose we have a population of fixed size N with two types, 1 and 2. Let n denote the number of type 1 alleles. One model for the evolution of this population is given by

$$\begin{aligned} \langle n + 1 | \mathbf{W} | n \rangle &= \frac{1}{\tau} \left[\frac{n}{N} \left(1 - \frac{n}{N}\right) (1 - u\tau) + \left(1 - \frac{n}{N}\right)^2 v\tau \right] \\ \langle n - 1 | \mathbf{W} | n \rangle &= \frac{1}{\tau} \left[\frac{n}{N} \left(1 - \frac{n}{N}\right) (1 - v\tau) + \left(\frac{n}{N}\right)^2 u\tau \right], \end{aligned} \quad (2.76)$$

where τ is a time scale such that $N\tau$ is the average generation time, u is the mutation rate from type 1 to type 2, and v is the reverse mutation rate. Since this is nonlinear, the solution to the master equation is very complex, but we can compute the steady state. Using Equation 2.11,

$$\begin{aligned}
\tilde{\pi}(n) &= \frac{1}{Z} \frac{\langle n|\mathbf{W}|n-1\rangle\langle n-1|\mathbf{W}|n-2\rangle\cdots\langle 1|\mathbf{W}|0\rangle}{\langle n-1|\mathbf{W}|n\rangle\langle n-2|\mathbf{W}|n-1\rangle\cdots\langle 0|\mathbf{W}|1\rangle} \\
&= \frac{1}{Z} \frac{\prod_{m=0}^{n-1} (m(N-m)(1-u\tau) + (N-m)^2v\tau)}{\prod_{m=1}^n (m(N-m)(1-v\tau) + m^2u\tau)} = \frac{1}{Z} \binom{N}{n} \frac{\prod_{m=0}^{n-1} (m(1-u\tau) + (N-m)v\tau)}{\prod_{m=1}^n ((N-m)(1-v\tau) + mu\tau)} \quad (2.77) \\
&= \frac{1}{Z} \binom{N}{n} \frac{\prod_{m=0}^{n-1} (Nv\tau + m(1-u\tau-v\tau))}{\prod_{m=1}^n (N(1-v\tau) - m(1-u\tau-v\tau))} = \frac{1}{Z} \binom{N}{n} \frac{\prod_{m=0}^{n-1} (\beta + m)}{\prod_{m=1}^n (\alpha + N - m)},
\end{aligned}$$

where

$$\alpha \equiv \frac{Nu\tau}{1-u\tau-v\tau}, \quad \beta \equiv \frac{Nv\tau}{1-u\tau-v\tau}. \quad (2.78)$$

We can now express this in terms of gamma functions:

$$\tilde{\pi}(n) = \frac{1}{Z} \binom{N}{n} \frac{\Gamma(n+\beta)\Gamma(N-n+\alpha)}{\Gamma(\beta)\Gamma(N+\alpha)}. \quad (2.79)$$

Solving for the normalization Z gives

$$Z = \sum_{n=0}^N \binom{N}{n} \frac{\Gamma(n+\beta)\Gamma(N-n+\alpha)}{\Gamma(\beta)\Gamma(N+\alpha)} = \frac{\Gamma(1-N-\alpha)\Gamma(1-\beta-\alpha)}{\Gamma(1-\alpha)\Gamma(1-N-\alpha-\beta)}. \quad (2.80)$$

Using a form of the Euler reflection identity ($\Gamma(z)\Gamma(1-z) = \Gamma(1-z-N)\Gamma(z+N)$ for $N \in \mathbb{Z}$),

$$\tilde{\pi}(n) = \frac{\Gamma(1-N-\alpha-\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(1-\alpha-\beta)} \binom{N}{n} \Gamma(n+\beta)\Gamma(N-n+\alpha). \quad (2.81)$$

2.4.6 Excited Electrons

Suppose we have n excited electrons (and n holes) in the valence and conduction bands. The valence band can hold N electrons and the conduction band M . Let $M \geq N$ so that $n = 0, 1, \dots, N$. This number drops by one when one of the n electrons meets one of the n holes, so that rate should be proportional to n^2 . The number increases by one when one of the $N-n$ valence electrons is excited into one of the $M-n$ holes, so that rate is proportional to $(N-n)(M-n)$. Therefore the rates are

$$\langle n+1|\mathbf{W}|n\rangle = u_n = \alpha(N-n)(M-n), \quad \langle n-1|\mathbf{W}|n\rangle = d_n = \beta n^2. \quad (2.82)$$

The steady state distribution is

$$\tilde{\pi}_n = \frac{1}{{}_2F_1(-M, -N; 1; \alpha/\beta)} \binom{N}{n} \binom{M}{n} \left(\frac{\alpha}{\beta}\right)^n, \quad (2.83)$$

where ${}_2F_1$ is the ordinary hypergeometric function. Note that this probably shouldn't be taken too seriously for n too close to N or M , when spatial effects will matter.

2.4.7 Two-Molecule Chemical Reaction

The band electron model is similar to a chemical process in which two molecules, A and B , can meet to form the molecule AB . Let N_A and N_B be the numbers of free A and B molecules, and they are sufficiently large that spatial diffusion can be neglected. Let n be the number of AB molecules. Then the rates are

$$\langle n+1|\mathbf{W}|n\rangle = u_n = \alpha(N_A - n)(N_B - n), \quad \langle n-1|\mathbf{W}|n\rangle = d_n = \beta n. \quad (2.84)$$

Assuming that $N_B \geq N_A$ (so $n = 0, 1, \dots, N_A$), the steady state distribution is

$$\tilde{\pi}_n = \frac{(-1)^{N_B}}{U(-N_B, 1 + N_A - N_B, -\beta/\alpha)} \binom{N_A}{n} \binom{N_B}{n} n! \left(\frac{\alpha}{\beta}\right)^{n+N_B}, \quad (2.85)$$

where U is the confluent hypergeometric function. Note that this model likely breaks down for n too close to N_A and N_B as in the previous example.

References

- [1] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, third edition, 2007.
- [2] F. S. Roberts. *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences*. Addison-Wesley, Reading, 1979.