Chapter 8
Markov Processes with General State Space

The focus of this chapter is that of discrete parameter Markov processes on a general (measurable) state space S. Some general considerations for the existence and uniqueness of invariant probabilities are provided.

Consider a discrete parameter stochastic process $\{X_n\}_{n\geq 0}$ with general state space S equipped with a σ-field $\mathcal{S}$ of subsets of S. Here the measurable space $(S, \mathcal{S})$ may be a countable set and $\mathcal{S}$ its power set as in the previous chapter, or more generally, S may be a metric space and $\mathcal{S}$ its (Borel) σ-field generated by the open subsets of S, for example. Unless stated otherwise, a map f on S into a metric space M is (implicitly) referred to as *measurable* (or more explicitly, *Borel measurable*) if it is measurable with respect to $\mathcal{S}$ on S and the Borel σ-field on M, i.e., if $f^{-1}(B) \in \mathcal{S}$ $\forall$ Borel sets B in M. Also, for any family $\{Y_\lambda : \lambda \in \Lambda\}$ of random variables on a probability space $(\Omega, \mathcal{F}, P)$, $\sigma\{Y_\lambda : \lambda \in \Lambda\}$ is the *σ-field generated by the family* i.e., it is the smallest σ-field $(\subset \mathcal{F})$ with respect to which Y_λ is measurable for each $\lambda \in \Lambda$.

Definition 8.1. A stochastic process $\{X_0, X_1, \ldots, X_n, \ldots\}$ having state space S equipped with a σ-field $\mathcal{S}$ has the *Markov property with regular transition probabilities* if for each $n \geq 0$,

$$P(X_{n+1} \in B | \sigma\{X_0, X_1, \ldots, X_n\}) = p_n(X_n, B), \quad B \in \mathcal{S}, \qquad (8.1)$$

where for each n,

1. For each $B \in \mathcal{B}$, $x \rightarrow p_n(x, B)$ is a measurable function on S.

2. For each $x \in S$, $B \rightarrow p_n(x, B)$ is a probability measure on $\mathcal{S}$.

A stochastic process having the Markov property is called a *Markov process* with transition probabilities $p_n(x, dy)$, $n \geq 0$. The Markov process is said to be *homogeneous* if the transition probabilities $p_n(x, dy)$ are the same for all $n = 1, 2, \ldots$, say $p(x, dy)$.

For the most part we will consider Markov processes with homogeneous transition probabilities, and unless explicitly stated otherwise, by a Markov process we will mean Markov process with homogeneous conditional probabilities from here out.

The special case in which there is a σ-finite measure ν on $(S, \mathcal{S})$ and a nonnegative measurable function $p(x, y)$ on $S \times S$ such that $\int_S p(x, y)\nu(dy) = 1$ and $p(x, B) = \int_B p(x, y)\nu(dy)$, for all $x \in S$, occurs often. In this case, (8.7) and (8.12) below are iterated integrals of $p(x_0, x_1)p(x_1, x_2) \ldots p(x_{n-1}, x_n)$. This was the case, for example, in Chapter 7, where S is finite or countable and ν is the counting measure.

The first task is to establish the following construction.

Proposition 8.1. Given an initial distribution μ and a transition probability $p(x, dy)$, there is a unique probability measure P_μ on the canonical space $(S^\infty, \mathcal{S}^{\otimes \infty})$ with the property (8.1) for the coordinate projections $X_n(\mathbf{x}) = x_n (n = 0, 1, 2, \ldots)$, where $\mathbf{x} = (x_0, x_1, x_2, \ldots) \in S^\infty \equiv S^{\{0,1,2,\ldots\}}$, and $\mathcal{S}^{\otimes \infty}$ is the product σ-field; i.e., the smallest σ-field on S^∞ for which each of the respective coordinate projection maps is measurable. Moreover,

$$P_\mu(X_{n+1} \in B_{n+1} | \sigma(X_0, \ldots, X_n)) = p(X_n, B_{n+1}), \quad B \in \mathcal{S}. \quad (8.2)$$

Proof. To begin, one constructs a collection of probability measures $P_{\mu,n}$, $n = 0, 1, 2, \ldots$, on the finite dimensional product spaces $(S^n, \mathcal{S}^{\otimes n})$, where $S^n = S \times \cdots \times S$, and $\mathcal{S}^{\otimes n} = \mathcal{S} \otimes \cdots \otimes \mathcal{S}$, ($n$-fold), to represent the respective distributions of $(X_0, X_1, \ldots, X_n)$, $n = 0, 1, 2 \ldots$.

To this end, for a bounded $\mathcal{S}^{\otimes n}$-measurable function f on S^n, define integrations on the product spaces iteratively, beginning with the innermost integral (with respect to $p(x_{n-1}, dx_n)$), keeping all variables except the last one, namely x_n, fixed; that is,

$$\int_S \cdots \int_S f(x_0, x_1, \ldots, x_{n-1}, x_n) p(x_{n-1}, dx_n) \cdots p(x_0, dx_1)\mu(dx_0)$$

$$= \int_S \cdots \int_S f_1(x_0, x_1, \ldots, x_{n-1}) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1)\mu(dx_0)$$

$$= \int_S \cdots \int_S f_2(x_0, x_1, \ldots, x_{n-2}) p(x_{n-3}, dx_{n-2}) \cdots p(x_0, dx_1)\mu(dx_0)$$

$$\vdots$$

$$= \int_S \int_S f_{n-1}(x_0, x_1) p(x_0, dx_1) \mu(dx_0)$$

$$= \int_S f_n(x_0) \mu(dx_0), \tag{8.3}$$

where

$$f_1(x_0, x_1, \ldots, x_{n-1}) = \int_S f(x_0, x_1, \ldots, x_{n-1}, x_n) p(x_{n-1}, dx_n),$$

$$f_2(x_0, x_1, \ldots, x_{n-2}) = \int_S f_1(x_0, x_1, \ldots, x_{n-1}) p(x_{n-2}, dx_{n-1}),$$

and so on. To justify this integration, first observe that $y \rightarrow f(x_0, \ldots, x_{n-1}, y)$ is $\mathcal{S}$-measurable for any fixed $(x_0, \ldots, x_{n-1}) \in S^n$; namely, it is clear for indicators of measurable rectangles $C = B_0 \times \cdots \times B_n, (B_i \in \mathcal{S}, i = 0, 1, \ldots, n)$, therefore, by the $\pi - \lambda$ theorem for all $C \in \mathcal{S}^{\otimes n}$. Since every bounded measurable function is a pointwise (uniform) limit of a sequence of simple functions, the measurability of $y \rightarrow f(x_0, \ldots, x_{n-1}, y)$ follows for all bounded $\mathcal{S}^{\otimes n}$-measurable functions f on S^n. The $\mathcal{S}^{\otimes n}$-measurability of the map $(x_0, \ldots, x_{n-1}) \rightarrow f_1(x_0, \ldots, x_{n-1}) = \int_S f(x_0, x_1, \ldots, x_{n-1}, y) p(x_{n-1}, dy)$ follows.

Now, to define $P_{\mu,n}$, take $f = \mathbf{1}_C, C \in \mathcal{S}^{\otimes n+1}$. Writing $C_{x_0, \ldots, x_{n-1}} = \{y \in S : (x_0, \ldots, x_{n-1}, y) \in C\}$, and $f_1(x_0, x_1, \ldots, x_{n-1}) = p(x_{n-1}, C_{x_0, \ldots, x_{n-1}})$, define

$$P_{\mu,0} = \mu, \ P_{\mu,1}(C) = \int_S \int_S \mathbf{1}_{C_{x_0}}(x_1) p(x_0, dx_1) \mu(dx_0) \ C \in \mathcal{S}^{\otimes 2}. \tag{8.4}$$

More generally, define for $n \geq 2$

$$P_{\mu,n}(C) = \int_S \cdots \int_S \mathbf{1}_{C_{x_0, \ldots, x_{n-1}}}(x_n) p(x_{n-1}, dx_n) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0)$$

$$= \int_S \cdots \int_S p(x_{n-1}, C_{x_0, \ldots, x_{n-1}}) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0). \tag{8.5}$$

In particular, for a measurable rectangle $C = B_0 \times \cdots \times B_n$, (8.5) reduces to

$$P_{\mu,1}(B_0 \times B_1) = \int_{B_0} p(x_0, B_1) \mu(dx_0),$$

and

$$P_{\mu,n}(B_0 \times \cdots \times B_n) = \int_{B_0} \cdots \int_{B_{n-1}} p(x_{n-1}, B_n) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0).$$

Finite additivity of $P_{\mu,n}$ follows by writing $\mathbf{1}_C = \sum_{j=1}^{m} \mathbf{1}_{C_j}$ for disjoint $C_j \in \mathcal{S}^{\otimes n}$, and countable additivity then follows by the monotone convergence theorem. This completes the construction of the finite dimensional distributions. It is simple to check that this collection of probability measures is *consistent* in the sense of the Kolmogorov existence theorem.[1] Thus, by the Kolmogorov existence theorem in the case that S is also a Borel subset of a Polish space or, more generally by Tulcea's theorem[2] for an arbitrary measurable space $(S, \mathcal{S})$, one arrives at a probability measure P_μ on $(S^\infty, \mathcal{S}^{\otimes\infty})$ such that the distribution of the projection $(X_0, X_1, \ldots, X_n) : S^\infty \to S^{n+1}$ is $P_{\mu,n}, n = 0, 1, 2, \ldots$. For this, let $B_{n+1} \in \mathcal{S}$. If $C \in \mathcal{S}^{\otimes(n+1)}$, then it follows from (8.5) that

$$P_{\mu,n+1}(C \times B_{n+1})$$
$$= \int_S \cdots \int_S \mathbf{1}_C(x_0, \ldots, x_n)\mathbf{1}_{B_{n+1}}(x_{n+1})P_{\mu,n+1}(dx_0 \times \cdots \times dx_{n+1})$$
$$= \int_{C \times S} p(x_n, B_{n+1})P_{\mu,n}(dx_0 \times \cdots \times dx_n). \tag{8.6}$$

If one takes $B_{n+1} = S$, then the consistency is checked: $P_{\mu,n+1}(C \times S) = P_{\mu,n}(C)$. This completes the construction of P_μ on $(S^\infty, \mathcal{S}^{\otimes\infty})$. Finally, note that (8.6) also expresses the Markov property in the form:

$$P_\mu(X_{n+1} \in B_{n+1}|\sigma(X_0, \ldots, X_n)) = p(X_n, B_{n+1}). \qquad \blacksquare$$

With the construction carried out in the proof of Proposition 8.1, given a transition probability $p(x, dy)$ and an initial distribution $\mu(dx)$, one may make the following definition.

Definition 8.2. A stochastic process $\{X_0, X_1, \ldots\}$ on an arbitrary probability space $(\Omega, \mathcal{F}, P)$ is Markov with transition probability $p(x, dy)$ and initial distribution $\mu(dx)$ if its distribution is P_μ on $(S^\infty, \mathcal{S}^{\otimes\infty})$. If $\mu = \delta_x$, then we write P_x for P_{δ_x}.

Note: To avoid a clutter of symbols, we will often abuse notation in probabilities associated with Markov processes $X = \{X_n : n = 0, 1, \ldots\}$ defined on a (possibly non-canonical probability space) $(\Omega, \mathcal{F}, P)$ as $P_\mu(X \in B)$ to indicate that X_0 has distribution μ. That is, we may use the expression for the corresponding probability in canonical space, where X_n is the nth coordinate projection on S^∞.

Proposition 8.2. If $\{X_n\}_{n \geq 0}^\infty$ has the Markov property, then one may obtain the distribution at $m \geq 1$ time points into the future inductively for $B_1, B_2, \ldots, B_m \in \mathcal{S}$, as

$$P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m|\sigma\{X_0, X_1, \ldots, X_n\})$$

---

[1] See BCPT p. 236 or Billingsley (1968), p. 235.
[2] see BCPT p. 168 or Billingsley (1986), pp. 510–511.

$$= P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m | \sigma\{X_n\})$$

$$= \int_{B_1} \cdots \int_{B_m} p(x_{m-1}, dx_m) \ldots p(x_1, dx_2) p(X_n, dx_1)$$

$$= P_x(X_1 \in B_1, \ldots, X_m \in B_m)|_{x=X_n}, \qquad (8.7)$$

where the integration is an iterated integral.

Proof. The first equality follows from the Markov property. The second follows from (8.6), as does the last. To prove this in detail, simply condition the left hand side of (8.7) on the larger σ-field $\sigma(X_0, \ldots, X_{n+m-1})$ and use the smoothing property of conditional expectations. Since

$$\int_{B_m} p(x_{m-1}, dx_m) = p(x_{m-1}, B_m),$$

the first integration yields

$$f_1(X_{n+1}, \ldots, X_{n+m-1}) = \mathbf{1}[X_{n+1} \in B_1, \ldots, X_{n+m-1} \in B_{m-1}] \cdot p(X_{m-1}, B_m),$$

in the notation introduced earlier. That is, the left side of (8.7) equals

$$\mathbb{E}(f_1(X_{n+1}, \ldots, X_{n+m-1}) | \sigma(X_0, \ldots, X_n)). \qquad (8.8)$$

Next taking the conditional expectation of (8.8), given $\sigma(X_0, \ldots, X_{n+m-2})$, one has

$$\mathbb{E}(f_2(X_{n+1}, \ldots, X_{n+m-2}) | \sigma(X_0, \ldots, X_n))$$

$$= \mathbf{1}[X_{n+1} \in B_1, \ldots, X_{n+m-2} \in B_{m-2}] \int_{B_{m-1}} p(X_{m-1}, B_m) p(X_{m-2}, dx_{m-1}).$$

Continuing in this way, the probability is ultimately given by a function of X_n of the form:

$$P(X_{n+m} \in B_m, \ldots, X_{n+1} \in B_1 | \sigma\{X_0, X_1, \ldots, X_n\})$$

$$= \mathbb{E}(f_{m-1}(X_{n+1}) | \sigma(X_0, \ldots, X_n))$$

$$= \mathbb{E} f_{m-1}(X_{n+1} | \sigma(X_n)) = \int_{B_1} f_{m-1}(x_1) p(X_n, dx_1). \qquad (8.9)$$

∎

Observe that taking $B_m = B \in \mathcal{S}$, $B_1 = B_2 = \cdots = B_{m-1} = S$ in (8.7), one has

$$P(X_{n+m} \in B | \sigma\{X_0, \ldots, X_n\}) = p^{(m)}(X_n, B), \qquad (8.10)$$

where the *m-step transition probabilities* are recursively given by

$$p^{(m+1)}(x, B) = \int_S p^{(m)}(y, B) p(x, dy), \quad B \in \mathcal{S}, \ x \in S, \tag{8.11}$$

with $p^{(1)}(x, B) \equiv p(x, B)$. Notice that by taking successive conditional expectations of $\mathbf{1}[X_0 \in B_0, \ldots, X_{n-1} \in B_{n-1}, X_n \in B_n]$ given $\sigma(X_0, \ldots, X_{n-1}), \sigma(X_0, \ldots, X_{n-2}), \ldots$, one obtains

$$P_\mu(X_0 \in B_0, \ldots, X_{n-1} \in B_{n-1}, X_n \in B_n)$$
$$= \int_{B_0} \int_{B_1} \cdots \int_{B_n} p(x_{n-1}, dx_n) p(x_{n-2}, dx_{n-1}) \cdots p(x_0, dx_1) \mu(dx_0)$$
$$= P_{\mu,n}(B_0 \times B_1 \times \cdots \times B_n),$$

where μ is the *initial distribution* of the process as defined by

$$\mu(B) = P(X_0 \in B), \quad B \in \mathcal{S}. \tag{8.12}$$

The following is another equivalent version of the Markov property that is commonly used.

Proposition 8.3. Let $\{X_n\}_{n \geq 1}$ be a Markov process with a transition probability $p(x, dy)$ and some initial distribution μ. Then the conditional distribution of the *after-n process* defined by $X_n^+ := (X_n, X_{n+1}, \ldots)$, given $\sigma\{X_0, \ldots, X_n\}$, is P_{X_n}. That is, it equals P_x on the event $[X_n = x] \subset \Omega$.

Proof. In view of (8.7), one has for finite dimensional events of the form $C = B_0 \times B_1 \times B_m \times S^\infty$, $B_i \in \mathcal{S}, 0 \leq i \leq m$, that $P(X_n^+ \in C | \sigma\{X_0, \ldots, X_n\}) = P_{X_n}((X_0, X_1, \ldots) \in C)$. Now observe that the collection of sets $C \in \mathcal{S}^{\otimes \infty}$ such that this equation holds is a λ-system, which contains a π-system of finite dimensional events. The assertion thus follows from an application of the $\pi - \lambda$ theorem. ∎

Situations in which there is an initial probability π for the Markov process which is invariant under the evolution are of particular interest, especially because when unique it represents the long term behavior of the process regardless of how it is initiated.

Definition 8.3. A probability π on $\mathcal{S}$ is said to be an *invariant probability* or *steady state distribution* for a Markov process $\{X_n\}_{n \geq 0}$ with transition probabilities $p(x, dy)$ if

$$\int_S p(x, B) \pi(dx) = \pi(B) \quad \text{for all } B \in \mathcal{S}. \tag{8.13}$$

Notice that if π is an invariant initial probability for $\{X_n\}_{n \geq 0}$, i.e., X_0 has the invariant distribution π, then the left side of (8.13) is $P(X_1 \in B)$. One has that $P(X_n \in B) = \pi(B)$, $B \in \mathcal{S}$, that is,

$$P(X_{n+1} \in B) = \int_S p(x, B) P(X_n \in dx) = \int_S p(x, B) \pi(dx) = \pi(B), \quad (8.14)$$

which for $n = 0$ is true by definition, and the general case follows by the Markov property and induction. In addition to questions of

 (i) *Existence*
(ii) *Uniqueness*

of invariant probabilities, one also seeks

(iii) *Basins of Attraction*

i.e., initial distributions under which convergence to a given invariant probability will hold, and

(iv) *Rates of Convergence*

to name a few of the central topics of the theory. In view of (8.7),

$$P(X_{n+1} \in B_1, \ldots, X_{n+m} \in B_m) = P(X_1 \in B_1, \ldots, X_m \in B_m), \quad B_i \in \mathcal{S}.$$
$$(8.15)$$

In particular, in this context the distribution of the after-n process $X_n^+ \equiv \{X_{n+m} : m = 0, 1, 2, \ldots\}$ coincides with that of $X_0^+ = \{X_m : m = 0, 1, 2, \ldots\}$ for each $n = 1, 2, \ldots$, a property referred to as *stationarity* of the process $\{X_0, X_1, \ldots\}$, as discussed in the earlier chapters of the text. From this perspective, theorems providing

 (v) *Law of Averages*

and

(vi) *Fluctuation Law*

in the forms of a strong law of large numbers (ergodic theorem) and a central limit theorem for averages of the form $\frac{1}{n} \sum_{j=0}^{n-1} f(X_j)$ in the presence of a unique invariant initial distributions are also essential to a complete theory.

For the next definition and elsewhere in the book, $B(S)$ denotes *the set of all real-valued bounded measurable functions on S*, equipped with the sup-norm $\| f \| = \sup_{x \in S} |f(x)|$.

Definition 8.4. Given a transition probability $p(x, dy)$ on $(S, \mathcal{S})$, the *transition operator T* is the map on $B(S)$ (into $B(S)$) defined by

$$Tf(x) = \int_S f(y) p(x, dy), \quad f \in B(S). \quad (8.16)$$

Note that the measurability of $x \to p(x, B)$ for every $B \in \mathcal{S}$ implies the measurability of Tf (Exercise 1). The transition operator is a positive linear contraction on $B(S)$ such that $T\mathbf{1} = \mathbf{1}$, where $\mathbf{1} \in B(S)$ is the constant function

on S with constant numerical value one (Exercise 1). In particular, given any such transition operator, one may recover $p(x, B) = T\mathbf{1}_B(x)$, $x \in S$, $B \in \mathcal{S}$.

Note that for a Markov process $\{X_n : n = 0, 1, 2, \dots\}$ with transition probability $p(x, dy)$, one has $Tf(x) = \mathbb{E}(f(X_1)|X_0 = x) \equiv \mathbb{E}(f(X_1)|\sigma(X_0))|_{X_0=x}$, which may be expressed as $\mathbb{E}_x f(X_1)$. In fact, by the Markov property and induction,

$$
\begin{aligned}
\mathbb{E}_x f(X_{n+1}) &= \mathbb{E}_x \mathbb{E}_x [f(X_{n+1}) \mid \sigma(X_0, \dots, X_n)] \\
&= \mathbb{E}_x \mathbb{E}_{X_n} f(X_{n+1}) \\
&= \mathbb{E}_x Tf(X_n) \\
&= \mathbb{E}_x \mathbb{E}(Tf(X_n)|\sigma(X_{n-1})) \\
&= \mathbb{E}_x TTf(X_{n-1}) = \mathbb{E}_x T^2 f(X_{n-1}) = \dots = \mathbb{E}T^n f(X_1) \\
&= T^{n+1} f(x), \quad x \in S, n \geq 0,
\end{aligned}
$$

i.e., T^n is the transition operator defined by the n-step transition probability $p^{(n)}(x, dy)$. The relation (8.13) implies (and is, therefore, equivalent to)

$$
\int_S (Tf)(x)\pi(dx) = \int_S f(x)\pi(dx) \quad \text{for all } f \in B(S). \tag{8.17}
$$

Thus (8.17) also defines the convenient notation "$\pi(dy) = \int_S p(x, dy)\pi(dx)$." Just as (8.13) implies (8.17), (8.14) implies

$$
\int_S (T^n f)(x)\pi(dx) = \int_S f(x)\pi(dx) \quad \text{for all } f \in B(S), \text{ for all } n \geq 1. \tag{8.18}
$$

One approach to obtain invariant probabilities is by consideration of long time steady state distributions. In particular, one might anticipate that if for some $x \in S$, the distribution $p^{(n)}(x, dy)$ of the state X_n converges weakly as $n \to \infty$ to some limit probability distribution π_x, then π_x should be invariant under continued evolution. However, as the following example shows, one must be careful.

Example 1 (Liggett). Let $S = \{0, 1, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots, \frac{m}{m+1}, \dots\}$, and let $\mathcal{S}$ be the power set of S. Then S is a metric space with the usual distance on the line, $d(x, y) = |x - y|$, and $\mathcal{S}$ is the Borel σ-field. Fix $0 < \theta \leq 1$ and define

$$
\begin{aligned}
p\left(0, \left\{\tfrac{1}{2}\right\}\right) &= \theta, & p\left(0, \left\{\tfrac{2}{3}\right\}\right) &= 1 - \theta \\
p\left(\tfrac{m}{m+1}, \left\{\tfrac{m+1}{m+2}\right\}\right) &= \theta, & p\left(\tfrac{m}{m+1}, \left\{\tfrac{m+2}{m+3}\right\}\right) &= 1 - \theta, \quad m = 1, 2, \dots. \\
p(1, \{0\}) &= \theta, & p\left(1, \left\{\tfrac{1}{2}\right\}\right) &= 1 - \theta.
\end{aligned}
$$

Then $p^{(n)}(x, dy)$ converges weakly to $\delta_{\{1\}}$, but this is clearly not an invariant probability.

Note that weak convergence requires convergence of the sequences of integrals $\int_S f(y) p^{(n+1)}(x, dy) = T^{n+1} f(x) = T^n(Tf)(x)$, for all $f \in C_b(S)$, as $n \to \infty$; recall $C_b(S) \subset B(S)$ denotes the subset of all bounded continuous functions on the metric space S. As usual whenever S is a metric space, we take $\mathcal{S}$ to be the Borel σ-field on S for the uniform norm on $C_b(S)$.

Definition 8.5. A transition probability $p(x, dy)$ on a metric space S is said to be *Feller continuous*, or *weakly Feller continuous*, if for every $f \in C_b(S)$, $Tf \in C_b(S)$. In this case one also says $p(x, dy)$ has the *(weak) Feller property*.

Observe that Feller continuity of $p(x, dy)$ means that the map $x \to p(x, dy)$, on S into the set $\mathcal{P}(S)$ of all probability measures on $(S, \mathcal{S})$, is *weakly continuous*. Moreover, T is a positive, linear contraction operator on $C_b(S)$ with $T1 = 1$. Conversely, if S is a compact metric space, then any such operator on $C_b(S)$ uniquely determines Feller transition probabilities $p(x, dy)$ by applying the Riesz Representation Theorem[3] from analysis to the bounded linear functional $f \to Tf(x)$, $f \in C_b(S)$, for each $x \in S$ (Exercise 6).

Notice that since $C_b(S)$ is measure-determining,[4] the condition (8.17) defining an invariant probability may be restricted to $f \in C_b(S)$.

Another obstacle to this approach to the determination of invariant probabilities is evident in the simple two-state example $p_{01} = p_{10} = 1$, $p_{00} = p_{11} = 0$. In this case, $\pi_0 = \pi_1 = 1/2$ is the unique invariant probability, but $p_{01}^{(n)}$ oscillates between 1 and 0 as a function of n. However, these oscillations can be averaged out by considerations of $\frac{1}{2n+1} \sum_{r=0}^{2n} p_{0j}^{(r)} \to 1/2$, as $n \to \infty$, for $j = 0, 1$. So this example suggests that time averaging may be required, and Liggett's Example 1 shows that the hypothesis of Feller continuity in Proposition 8.4 cannot in general be dispensed with in the weak convergence approach to invariant probabilities (Exercise 5). From a probabilistic perspective, note also that

$$\frac{1}{m} \sum_{r=0}^{m-1} p^{(r)}(x, B) = \frac{1}{m} \sum_{r=0}^{m-1} \mathbb{E}_x \mathbf{1}[X_r \in B] = \mathbb{E}_x \left(\frac{\sum_{r=0}^{m-1} \mathbf{1}[X_r \in B]}{m} \right) \qquad (8.19)$$

is the expected proportion of visits to the set $B \in \mathcal{S}$ in time 0 to $m - 1$, starting from x.

Proposition 8.4. Suppose S is a metric space and $p(x, dy)$ is a Feller continuous transition probability on $(S, \mathcal{S})$. (a) If for some $x \in S$ there is a sequence of integers $1 \leq n_1 < n_2 < \dots$ such that, as $k \to \infty$,

$$\frac{1}{n_k} \sum_{r=0}^{n_k-1} p^{(r)}(x, dy) \text{ converges weakly to } \pi_x(dy) \qquad (8.20)$$

---

[3] BCPT p 237.
[4] See BCPT, p. 11.

for some probability measure π_x, then π_x is an invariant for $p(x, dy)$.

(b) If, for some sequence $1 \leq n_1 < n_2 \ldots$, (8.20) holds for every $x \in S$ with the same limit $\pi_x = \pi$ for all x, then π is the unique invariant probability.

Proof.

(a) The relation (8.20) says that

$$\frac{1}{n_k} \sum_{r=0}^{n_k-1} (T^r f)(x) \longrightarrow \int_S f(y) \pi_x(dy) \quad \text{for all } f \in C_b(S). \qquad (8.21)$$

Replacing f by Tf (which belongs to $C_b(S)$ by hypothesis), one gets

$$\frac{1}{n_k} \sum_{r=1}^{n_k} T^r f(x) \longrightarrow \int_S Tf(y) \pi_x(dy) \quad \text{for all } f \in C_b(S). \qquad (8.22)$$

But the difference between the left sides of (8.21) and (8.22) equals in magnitude $|(T^{n_k} f)(x) - f(x)|/n_k \leq 2 \sup\{|f(x)| : x \in S\}/n_k$, which goes to zero as $k \to \infty$. Hence the limits in (8.21) and (8.22) are the same. Thus, since $C_b(S)$ is measure-determining, one has $\pi_x(dz) = \int_S p(y, dz)\pi_x(dy)$; see Lemma 1 below.

(b) By (a), π is invariant. Suppose that, under the hypothesis of part (b), π' is another invariant probability, and then integrating the two sides of (8.21) with respect to π', one obtains

$$\frac{1}{n_k} \sum_{r=0}^{n_k-1} \int_S (T^r f)(x) \pi'(dx) \longrightarrow \int_S \left[\int_S f(y)\pi(dy) \right] \pi'(dx). \qquad (8.23)$$

By invariance of π', the left side equals $\int f d\pi'$ (see (8.18)), while the right side is $\int f d\pi$. Thus $\int f d\pi' = \int f d\pi$ for every $f \in C_b(S)$, implying $\pi' = \pi$ since $C_b(S)$ is measure-determining. ∎

Lemma 1. If Q_1 and Q_2 are probability measures on the Borel σ-field of a metric space S such that $\int_S f dQ_1 = \int_S f dQ_2$ for all bounded continuous real-valued functions f on S, then $Q_1 = Q_2$.

Proof. Let $\mathcal{C}$ be the collection of Borel sets B such that $Q_1(B) = Q_2(B)$. Then it is simple to check that $\mathcal{C}$ is a σ-field. Since $\mathcal{B}$ the Borel σ-field is the smallest σ-field containing all closed sets, it is sufficient to show that $\mathcal{C}$ contains all closed sets. For this, it is enough to show that for each (closed) $F \subset S$, there exists a sequence of nonnegative functions $\{f_n\} \subset C_b(S)$ such that $f_n \downarrow \mathbf{1}_F$ as $n \uparrow \infty$. Since F is closed, one may view $x \in F$ in terms of the equivalent condition that $\rho(x, F) = 0$, where $\rho(x, F) := \inf\{\rho(x, y) : y \in F\}$. Let $h_n(r) = 1 - nr$ for $0 \leq r \leq 1/n$, $h_n(r) = 0$ for $r \geq 1/n$. Then take $f_n(x) = h_n(\rho(x, F))$. In

particular, $\mathbf{1}_F(x) = \lim_n f_n(x), x \in S$, and Lebesgue's monotone convergence theorem applies to show $F \in \mathcal{C}$.   ∎

As an immediate corollary, one gets the following corollary:

Corollary 8.5. If a transition probability $p(x, dy)$ on a metric space has the (weak) Feller property and there exists a Caesaro limit in the weak topology, namely,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{r=0}^{n-1} p^{(r)}(x, dy) = \pi(dy) \tag{8.24}$$

such that the probability measure π does not depend on x, then π is the unique invariant probability.

Many important Markov processes do not admit an invariant probability, such as the case, for example, of a random walk on $\mathbb{R}^k$ with an arbitrary step size distribution $Q \neq \delta_{\{0\}}$ (Exercise 8). There is one case, namely that of a compact state space, where every Feller transition probability admits at least one invariant probability.

Proposition 8.6. Let S be a compact metric space and $\mathcal{S}$ its Borel σ-field. If $p(x, dy)$ is a Feller transition probability on $(S, \mathcal{S})$, then it admits an invariant probability.

Proof. Fix $x \in S$ and consider the sequence of probability measures $\mu_n, n \geq 1$, given by $\mu_n(B) = (1/n) \sum_{m=1}^{n} p^{(m)}(x, B), B \in \mathcal{S}$. Since $\mathcal{P}(S)$ is a weakly compact metric space,[5] there exists a subsequence $\{\mu_{n_k} : k = 1, 2, \dots\}$, which converges weakly to a probability measure π_x, say. By Proposition 8.4, π_x is invariant.   ∎

As a simple corollary, we get the following result for finite Markov chains.

Corollary 8.7. A Markov chain on a finite state space S has at least one invariant probability.

Proof. This follows from Proposition 8.6 by making S a compact metric space with the metric $d(x, y) = 1$ if $x \neq y$, $d(x, x) = 0$. Then $\mathcal{S} \equiv \mathcal{B}(S)$ is the class of all subsets of S, and every real-valued function on S is continuous.   ∎

Remark 8.1. A direct proof of Corollary 8.7, which does not use Proposition 8.6 will be given later (see Corollary13.8 in Chapter 13).

The next approach to invariant probabilities is based on *symmetries*. For the definition below, consider a stationary Markov process $\{X_n\}_{n \geq 0}$ on a state space $(S, \mathcal{S})$ with transition probability $p(x, dy)$. Since the distribution of such a process is invariant under time shift, i.e., $\{X_n\}_{n \geq 0}$ and $\{X_n\}_{n \geq k}$ have the same distribution, one may use Kolmogorov's existence theorem to construct a stationary Markov

---

[5] See BCPT, p.142.

process $\{Z_n\}_{-\infty<n<\infty}$ having the same transition probability p and the same invariant probability π.

Definition 8.6. A Markov process with a transition probability p and an invariant probability π is said to be *time-reversible* if the stationary Markov process $\{Z_n\}_{-\infty<n<\infty}$, with this transition probability and this invariant distribution, has the same distribution as the *time-reversed* process $\{Y_n\}_{-\infty<n<\infty}$, where $Y_n := Z_{-n}$ $(-\infty < n < \infty)$. We refer to $\{Z_n\}_{-\infty<n<\infty}$ as the *double-sided version*.

In the context of the "movie metaphor," the statistics of a stationary data stream does not depend on when viewing begins, while a time-reversible data stream sequence is the same whether it is viewed forward or backward.

For the propositions below, assume that the transition probability $p(x, dy)$ has a density $p(x, y)$ with respect to a σ-finite measure μ on $(S, \mathcal{S})$, with $(x, y) \to p(x, y)$ measurable (on $(S \times S, \mathcal{S} \otimes \mathcal{S})$ into $([0, \infty), \mathcal{B}_{[0,\infty)})$).

Proposition 8.8 *(Detailed Balance Condition).* Let $\pi(dy)$ be a probability measure on $(S, \mathcal{S})$ with a density $\pi(y)$ with respect to μ. (a) If

$$\pi(x)p(x, y) = \pi(y)p(y, x) \quad \text{a.e. } (\mu \times \mu), \tag{8.25}$$

then π is a time-reversible invariant probability for the Markov process. (b) For a Markov process with transition probability density $p(x, y)$ and invariant probability density $\pi(y)$, (8.25) is necessary for the process to be time-reversible.

Proof.

(a) Let p and π satisfy (8.25). Then for every Borel measurable f,

$$\begin{aligned}
\int_S Tf(x)\pi(dx) &= \int_S Tf(x)\pi(x)\mu(dx) = \int_S \left(\int_S f(y)p(x, y)\mu(dy) \right) \pi(x)\mu(dx) \\
&= \int_S \int_S f(y)p(x, y)\pi(x)\mu(dy)\mu(dx) \\
&= \int_S \left(\int_S f(y)p(y, x)\mu(dx) \right) \pi(y)\mu(dy) \\
&= \int_S f(y)\pi(y)\mu(dy) = \int_S f(y)\pi(dy),
\end{aligned}$$

implying π is invariant. Let $\{X_n\}_{n\geq 0}$ be a stationary Markov process with transition probability p and invariant initial distribution π. Then, by the Markov property, the joint density of $(X_n, X_{n+1}, \ldots, X_{n+k})$, with respect to $\mu \times \cdots \times \mu$, at $(y_0, y_1, \ldots, y_k) \in S^{k+1}$ is

$$g(y_0, y_1, \ldots, y_k) := \pi(y_0)p(y_0, y_1)p(y_1, y_2) \ldots p(y_{k-1}, y_k), \tag{8.26}$$

while the joint density of $(X_{n+k}, X_{n+k-1}, \ldots, X_n)$, at the same point $(y_0, y_1, \ldots, y_k) \in S^{k+1}$, is

$$h(y_0, y_1, \ldots, y_k) := g(y_k, y_{k-1}, \ldots, y_1, y_0)$$

$$= \pi(y_k)p(y_k, y_{k-1})p(y_{k-1}, y_{k-2}) \cdots p(y_1, y_0)$$

$$= \pi(y_{k-1})p(y_{k-1}, y_k)p(y_{k-1}, y_{k-2}) \cdots p(y_1, y_0)$$

$$= p(y_{k-1}, y_k)\pi(y_{k-1})p(y_{k-1}, y_{k-2}) \cdots p(y_1, y_0)$$

$$= p(y_{k-1}, y_k)\pi(y_{k-2})p(y_{k-2}, y_{k-1}) \cdots p(y_1, y_0)$$

$$\vdots$$

$$= p(y_{k-1}, y_k)p(y_{k-2}, y_{k-1}) \cdots p(y_1, y_2)\pi(y_1)p(y_1, y_0)$$

$$= p(y_{k-1}, y_k)p(y_{k-2}, y_{k-1}) \cdots p(y_1, y_2)\pi(y_0)p(y_0, y_1)$$

$$= g(y_0, y_1, \ldots, y_k).$$

Since this is true for all $k \geq 1$, the finite dimensional distributions of the double-sided version $\{Z_n\}_{-\infty<n<\infty}$ and $\{Y_n\}_{-\infty<n<\infty}$ with $Y_n := Z_{-n}$ $(-\infty < n < \infty)$, described in Definition 8.6, coincide. Thus, using the $\pi - \lambda$ theorem,[6] it follows that the process $\{Z_n\}_{-\infty<n<\infty}$ and its time-reversal $\{Y_n\}_{-\infty<n<\infty}$ have the same distribution.

(b) For the stationary Markov process $\{Z_n\}_{-\infty<n<\infty}$ to be time-reversible, it is necessary that the distribution of (Z_0, Z_1) is the same as that of $(Y_0, Y_1) \equiv (Z_0, Z_{-1})$. But the latter has the same distribution as (Z_1, Z_0). The left side of (8.25) is the p.d.f. of (Z_0, Z_1) at (x, y), while the right side is the p.d.f. of (Z_1, Z_0) at (x, y). Thus for time-reversibility, (8.25) must hold. ■

If π is an invariant probability, then by Jensen's inequality,

$$\int_S \left(\int_S |f(y)|p(x, dy) \right)^2 \pi(dx) \leq \int_S \int_S f^2(y)p(x, dy)\pi(dx) = \int_S f^2(y)\pi(dy),$$

so that one may extend the transition operator T to $L^2(S, \pi) \supseteq B(S)$. We will now show that, in analytical terms, time-reversibility of a Markov process means that the transition operator T is *self-adjoint* on $L^2(S, \pi)$, i.e.,

$$\langle Tf, g \rangle = \langle f, Tg \rangle \qquad \text{for all } f, g \in L^2(S, \pi). \tag{8.27}$$

Here $\langle \;\; \rangle$ is the *inner product* on the Hilbert space $L^2(S, \pi)$,

$$\langle g, h \rangle = \int_S g(y)h(y)\pi(dy).$$

We will denote by $\|g\|$ the L^2–*norm*: $\|g\|^2 = \langle g, g \rangle$.

Proposition 8.9. Let π be an invariant probability of a Markov process. (a) The transition operator T is a contraction on $L^2(S, \pi)$. (b) If π is a time-reversible invariant probability, then T is self-adjoint.

---

[6] See BCPT, p. 4.

Proof.

(a) This is proved in (8.28).

(b) Let $f, g \in L^2(S, \pi)$, and assume π is time-reversible in the sense of Definition 8.6. Then, conditioning on X_0, one has

$$
\begin{aligned}
\langle Tf, g \rangle &= \int_S \left(\int_S f(y) p(x, dy) \right) g(x) \mu(dx) \\
&= \mathbb{E}\big(\mathbb{E}(f(X_1) | \sigma(X_0)) g(X_0)\big) = \mathbb{E} f(X_1) g(X_0) \\
&= \mathbb{E} f(X_0) g(X_1) = \langle f, Tg \rangle.
\end{aligned}
$$
∎

Recall in the case of finite S that, given an initial distribution μ for X_0, the distribution μ_1 of X_1 may be obtained by the transformation $\mu \to \mu_1 = \mathbf{p}' \mu$, where $\mathbf{p}'$ is the transpose matrix, see (7.12). More generally, one may define an adjoint operator as follows.

Definition 8.7. Given a transition probability $p(x, dy)$ on $(S, \mathcal{S})$, the *adjoint* linear operator T^* is defined on the linear space $\mathcal{M}(S)$ of all finite signed measures on $(S, \mathcal{S})$ by

$$
(T^* \mu)(B) = \int_S p(x, B) \mu(dx) \qquad (B \in \mathcal{S}, \mu \in \mathcal{M}(S)). \tag{8.28}
$$

In general, if μ is a probability measure, then $T^* \mu$ is the distribution of X_1 where X_0 has distribution μ. In particular, π is an invariant probability if and only if

$$
T^* \pi = \pi. \tag{8.29}
$$

To see the connection between the $L^2(S, \pi)$-adjoint of T and this more general operator, then, irrespective of (8.27), identify $f \in L^2(S, \pi)$ with the signed measure $f d\pi$, and note that $T^*(f d\pi)(dy)$ is given by

$$
\int_S g(y) T^*(f d\pi)(dy) = \int_S \int_S g(y) p(x, dy) f(x) \pi(dx) = \int_S Tg(x) f(x) \pi(dx)
$$
$$
= \langle Tg, f \rangle = \langle g, T^* f \rangle, g \in L^2(S, \pi), \tag{8.30}
$$

where, by an obvious abuse of notation, $T^* f \in L^2(S, \pi)$ is given by the $L^2(S, \pi)$-adjoint operator to T. In the interpretation of T^* as an operator on $L^2(S, \pi)$, 1 is an *eigenvalue* of T^* with the constant *eigenvector* $f(\cdot) \equiv 1$. For the adjoint operator on $\mathcal{M}(S)$, this eigenvector corresponds to the invariant measure $\pi(dy) = 1 \cdot \pi(dy)$.

Define $T^{*n} \mu = T^*(T^{*n-1} \mu)$ iteratively on $\mathcal{M}(S)$. In the case μ is a probability measure, and X_0 has distribution μ, $T^* \mu$ is the distribution of X_1, $T^{*2} \mu$ is the distribution of $X_2, \ldots, T^{*n} \mu$ is the distribution of X_n. Thus, whereas iterates of the transition operator T govern the "evolution of states" via $T^n f(x) = \mathbb{E}_x f(X_n)$, the

iterates of the adjoint T^* govern the "evolution of probability distributions" of the Markov process.

Now, T^{*n} is a linear operator on $\mathcal{M}(S)$, as is T^n on $C_b(S)$. The term *adjoint operator* given to T^* (or, T^{*n}) is more fully justified in terms of the basic identities

$$\int_S Tf(x)\mu(dx) = \int_S f(x)(T^*\mu)(dx), \quad \int_S T^n f(x)\mu(dx) = \int_S f(x)(T^{*n}\mu)(dx).$$
(8.31)

The first equality in (8.31) follows from (8.28), first for simple functions f and then by approximating $f \in B(S)$ uniformly by simple functions. The second equality in (8.31) follows by induction on n. When μ is a probability measure, then the second equality says $\mathbb{E}_\mu[\mathbb{E}(f(X_n) \mid X_0)] = \mathbb{E}_\mu f(X_n)$, with X_0 having distribution μ.

Example 2. $S = \{0, 1\}$, $\mathbf{p} = \begin{bmatrix} a & 1-a \\ 1-b & b \end{bmatrix}$, $0 \le a, b \le 1$. S is a metric space with the discrete metric $d(0, 0) = d(1, 1) = 0$, $d(1, 0) = d(0, 1) = 1$, $\mathcal{S}$ is the power set, and every function $f : S \to \mathbb{R}$ is a bounded, continuous function. By the computation at the end of the previous chapter, one has for $a + b < 2$, with $a + b \ne 0$,

$$\lim_{n \to \infty} \mathbf{p}^{(n)} = \begin{bmatrix} \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \\ \frac{1-b}{2-a-b} & \frac{1-a}{2-a-b} \end{bmatrix}.$$

In particular, the invariant probability (vector) $\pi = (\pi_0, \pi_1)'$ is given by

$$\begin{aligned} \pi_0 &= \lim_{n\to\infty} p_{i0}^{(n)} = \tfrac{1-b}{2-a-b} \\ \pi_1 &= \lim_{n\to\infty} p_{i1}^{(n)} = \tfrac{1-a}{2-a-b} \,, \qquad i = 0, 1. \end{aligned}$$

Alternatively, one may determine π_0, π_1 from time-reversibility via the detailed balance and total probability one equations. In this case one could then use the $L^2(S, \pi)$ theory for self-adjoint operators to obtain convergence (Exercise 7).

In the cases $a + b = 2$ and $a + b = 0$, one has $\mathbf{p} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\mathbf{p} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, respectively. In the former case every probability on $S = \{0, 1\}$ is an invariant probability, and hence there are infinitely many invariant probabilities. In the latter case there is a unique invariant probability given by $\pi_0 = \pi_1 = \frac{1}{2}$; but $\mathbf{p}^n$ does not have a limit, although (8.24) holds.

Example 3. Suppose $S = \mathbb{R}$ with Borel σ-field $\mathcal{S}$. Let $\varepsilon_1, \varepsilon_2, \ldots$ be an i.i.d. sequence of standard normal random variables and let $b \in \mathbb{R}$. Consider the sequence of random variables

$$X_{n+1} = bX_n + \varepsilon_{n+1}, \quad n = 0, 1, 2, \ldots .$$

Iterating the recursion, one has that

$$X_{n+1} = b^{n+1} X_0 + \sum_{j=0}^{n} b^j \varepsilon_{n+1-j}.$$

In particular, the m-step transition probability $p^{(m)}(x, dy)$ is given by the Gaussian distribution with mean $b^m x$ and variance $\sum_{j=0}^{m-1} b^{2j} = \frac{b^{2m}-1}{b^2-1}$ if $|b| \neq 1$. In the case $b = \pm 1$, $p^{(m)}(x, dy)$ is Gaussian with mean $(\pm 1)^m x$ and variance m. In any case $p(x, dy)$ clearly has the (weak) Feller property, and in particular, if $|b| < 1$, then $p^{(m)}(x, dy)$ converges (weakly) as $m \to \infty$ to the invariant distribution $\pi(dy) = \frac{1}{\sqrt{2\pi(1-b^2)^{-1}}} e^{-\frac{1-b^2}{2} y^2} dy$, $y \in S = (-\infty, \infty)$, possessing a Gaussian density with respect to Lebesgue measure. Note that for $|b| < 1$, if X is $N(0, \sigma^2)$ and Z is standard normal and independent of X, then $X =^{\text{dist}} bX + Z$ (equality in distribution) if and only if $\sigma^2 = (1 - b^2)^{-1}$. One may check that π is also a time-reversible invariant probability (Exercise 13).

Example 4 *(Random Walk on a Finite Graph).*   A finite graph consists of a finite set $S = \{v_1, \ldots, v_k\}$ of k *vertices* together with a relation $\mathcal{E} \subset \{1, \ldots, k\} \times \{1, \ldots, k\}$, with the property that $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$, defining *edges* as follows: there is an edge e_{ij} connecting vertices v_i and v_j if and only if $(i, j) \in \mathcal{E}$, denoted by means of the obvious abuse of notation $e_{ij} \in \mathcal{E}$. The graph is said to be *connected* if for any pair of distinct vertices v_i, v_j there is a path of $m \geq 1$ edges $e_{ii_1}, e_{i_1 i_2}, \ldots e_{i_{m-1} i_m}$ with $i_m = j$. For a fixed vertex v_i, the integer $d_i := \text{card}\{j : e_{ij} \in \mathcal{E}\}$ is called the *degree* of v_i. A *random walk on a finite connected graph* $(S, \mathcal{E})$ may be defined as a Markov chain with state space S and transition probabilities given by $p_{v_i, v_j} = 1/d_i$ if and only if $e_{ij} \in \mathcal{E}$, else $p_{v_i, v_j} = 0$. It is straightforward to check that up to normalization, the vertex degrees define the unique time-reversible invariant probability for a random walk on a finite connected graph (see Exercise 12).

An approach to the construction of invariant probabilities, similar to that in Proposition 8.4 but which is valid without the Feller property, is given below.

Proposition 8.10.   Let $p(x, dy)$ be a transition probability on a state space $(S, \mathcal{S})$.

a If for some $x \in S$ there is a sequence of integers $1 \leq n_1 < n_2 < \cdots$ such that, as $k \to \infty$,

$$\frac{1}{n_k} \sum_{r=0}^{n_k-1} p^{(r)}(x, B) \longrightarrow \pi_x(B) \quad \text{for all } B \in \mathcal{S} \tag{8.32}$$

for some probability measure π_x on $(S, \mathcal{S})$, then π_x is an invariant probability for p.

b If, for some sequence $\{n_k : k \geq 1\}$, (8.32) holds for every $x \in S$ with the same limit $\pi_x = \pi$ for all x, then π is the unique invariant probability for p.

Proof. The proof is essentially the same as that of Proposition 8.4, with $C_b(S)$ replaced by $B(S)$—the space of real-valued bounded measurable functions on S. Note that (8.32) implies

$$\frac{1}{n_k} \sum_{r=0}^{n_k-1} (T^r f)(x) \longrightarrow \int_S f(y)\pi_x(dy) \quad \text{for all } f \in B(S). \qquad \blacksquare$$

To close this chapter, let us note that a large class of examples of Markov chains occur as functions of a given, perhaps more primitive Markov chain. Of course, one-to-one functions are simply relabeling of the states and do not affect the dependence structure. A more general class of functions can be obtained as follows[7].

Definition 8.8. A measurable function φ on $(S, \mathcal{S})$ to a measurable space $(S', \mathcal{S}')$, is said to be an invariant function of a group G of transformations on S if (i) $\varphi(gx) = \varphi(x)$ for all $g \in G$, $x \in S$. If, in addition, (ii) every measurable invariant function is a measurable function of φ, then φ is said to be a maximal invariant.

Example 5.

1. For each $x \in S$, the *orbit* of x under G is defined by $o(x) = \{gx : g \in G\}$. Note that each invariant function is constant on orbits. Let S and S' be metric spaces and $\varphi : S \to S'$ a measurable surjection such that (a) φ is constant on orbits, (b) $\varphi(x) \neq \varphi(y)$ if $o(x) \neq o(y)$, i.e., φ is a relabeling of $o(x)$, and S' may be viewed as a relabeling of the space of orbits. Then φ is a maximal invariant since (i) invariance is obvious by (a), and (ii) if $\varphi(x) = \varphi(y)$, then $\rho(x) = \rho(y)$ for any invariant function ρ by (b), i.e., ρ is a function of φ.
2. $\varphi(x) = |x|$ is a maximal invariant of the reflection group $\{e, -e\}$, where $ex = x$, $(-e)x = -x$, $x \in S = \mathbb{R}$, $S' = [0, \infty)$.

Proposition 8.11. Suppose $X = \{X_n\}$ is a Markov process on S whose transition probabilities are invariant under the group G of transformations from S to S', i.e.,

$$p(gx, g(B)) = p(x, B), \quad \forall\, x \in S, g \in G, B \in \mathcal{S}.$$

If φ is a maximal invariant, then $\{\varphi(X_n)\}$ is Markov.

Proof. Take conditional expectations with respect to the larger $\sigma(X_m : m \leq n)$, followed by the smaller $\sigma(\varphi(X_m) : m \leq n), n = 1, 2, \ldots$, to get

$$P(\varphi(X_{n+1}) \in B | \varphi(X_m), m \leq n) = \mathbb{E}\{p(X_n, \varphi^{-1}(B)) | \varphi(X_m), m \leq n\}. \qquad (8.33)$$

By invariance of φ, $\varphi \circ g = \varphi$, and one has from (i),

---

[7] The continuous parameter version of this result is given in Bhattacharya and Waymire (1990, 2009), pp. 502–503, for non-injective functions of a Markov process. Also see Bhattacharya and Waymire (1990).

$$p(x, \varphi^{-1}(B)) = p(g^{-1}x, g^{-1}(\varphi^{-1}(B)))$$
$$= p(g^{-1}x, \varphi^{-1}(B)). \tag{8.34}$$

That is, the function $x \to p(x, \varphi^{-1}(B))$ is invariant. By (ii), therefore, it is a function $q(\varphi(x), B)$, say, of φ. Thus,

$$P(\varphi(X_{n+1}) \in B | \varphi(X_m), m \le n) = \mathbb{E}\{p(X_n, \varphi^{-1}(B)) | \varphi(X_m), m \le n\}$$
$$= \mathbb{E}\{q(\varphi(X_n), B) | \varphi(X_m), m \le n\}$$
$$= q(\varphi(X_n), B). \tag{8.35}$$

Thus, $\{\varphi(X_n)\}$ is Markov with one-step transition probabilities $q(\varphi(x), B)$.   ∎

Example 6 *(Reflecting Simple Symmetric Random Walk).* Consider the unrestricted simple symmetric random walk on $\mathbb{Z}$ starting at the origin, defined by $S_n := \sum_{j=1}^{n} X_j, n \ge 1,\ S_0 = 0$, where the displacements $X_n : n \ge 1$ are i.i.d. ± 1-valued symmetric Bernoulli random variables. Then, since $p_{ij} = \frac{1}{2}\delta_{i-1}(j) + \frac{1}{2}\delta_{i+1}(j), i, j \in \mathbb{Z}$ is invariant under the reflection group $G = \{e, -e\}$ where $e(i) = i, i \in \mathbb{Z}$, it follows that $\{R_n := |S_n|\}_n$ is a Markov chain. (Also see Example 11.1.)

Exercises

1. Prove that for a transition probability p the measurability $x \to p(x, B)$ for all $B \in \mathcal{S}$ implies that $x \to Tf(x)$ is measurable for all $f \in \mathbb{B}(S)$. Show that (i) T is a linear operator on $\mathbb{B}(S)$, (ii) $||Tf|| \le ||f||, f \in \mathbb{B}(S), ||f|| = \sup_{x \in S} |f(x)|$, (iii)$T\mathbf{1} = \mathbf{1}$, where $\mathbf{1}(x) = 1$, for all $x \in S$, and (iv) $Tf \ge 0$ on S if $f \in \mathbb{B}(S)$ is a nonnegative function.
2. Let $p(x, dy)$ be a transition probability on $(S, \mathcal{S})$.

    (a) Show that $x \to P_x(B)$ is $\mathcal{S}$-measurable for all $B \in \mathcal{S}^{\otimes\infty}$, and letting $\mathbf{X}$ denote the identity map on S^∞, the function $y \to \mathbb{E}_y f(\mathbf{X})$ is $\mathcal{S}$-measurable for all bounded measurable $f : S^\infty \to \mathbb{R}$.
    (b) For every bounded $\mathcal{S}^{\otimes n}$-measurable function f on S^n, show that the function $(x_0, x_1, \ldots, x_{n-1}) \to \int_S f(x_0, x_1, \ldots, x_{n-1}, y)p(x_{n-1}, dy)$ is $\mathcal{S}^{\otimes n}$-measurable.

3. Express time-reversibility and detailed balance without requiring densities. [*Hint*: Detailed balance may be stated as $\pi(dx)p(x, dy) = \pi(dy)p(y, dx)$, suitably interpreted, and the consequent time-reversibility compares the joint distribution of $(X_0, \ldots, X_k)$ with that of $(X_k, X_{k-1}, \ldots, X_0)$.]
4. Prove that $T^n f(x) = \int_S f(y)p^{(n)}(x, dy) = \mathbb{E}(f(X_n)|X_0 = x), x \in S$, for all bounded, measurable functions f on S, and $n \ge 1$.

5. Show that the transition probability p in Example 1 (i) is not Feller continuous, and (ii) does not satisfy the hypothesis of Proposition 8.10. [*Hint*: (i) has the relative topology of $[0, 1]$ so that all points $m/(m + 1)$, $m = 0, 1, \ldots$, are isolated and only 1 is a point of accumulation. Hence a real-valued function on S is continuous if and only if $f(\frac{m}{m+1}) \to f(1)$ as $m \to \infty$. For (ii), $p^{(n)}(x, dy) \to \delta(dy)$ weakly as $n \to \infty$ for all $x \in S$.]

6. Let S be a locally compact separable metric space. Suppose that T is a positive linear contraction operator on $C_b(S)$ with $T1 = 1$. Use the Riesz Representation Theorem to show that T uniquely determines Feller transition probabilities $p(x, dy)$ such that $Tf(x) = \int_S f(y) p(x, dy)$.

7. In the case of the two-state Markov chain in Example 2, use time-reversibility to compute the invariant probability π and establish convergence by an appeal to the spectral theorem for self-adjoint linear operators on $L^2(S, \pi)$.

8. Let $X_0, X_n := X_0 + Z_1 + \cdots + Z_n$ $(n \geq 1)$ be a (general) random walk $S = \mathbb{R}^k$ with step size distribution Q, i.e., $\{Z_n : n \geq 1\}$ is an i.i.d. sequence with common distribution Q on $(\mathbb{R}^k, \mathcal{B}^k)$, independent of X_0. Show that (i) $\{X_n : n \geq 0\}$ is Markov and (ii) no invariant probability exists if $Q \neq \delta_{\{0\}}$.

9. (*Birth–Death Chain with Two Reflecting Boundaries*) Let $S = \{0, 1, 2, \ldots, d\}$ $(d > 1)$, $p(x, x + 1) = \beta_x$, $p(x, x - 1) = \delta_x \equiv 1 - \beta_x$, with $0 < \beta_x < 1$ $(x = 1, 2, \ldots, d - 1)$, $\beta_0 \equiv p(0, 1) = 1$, $\delta_d \equiv p(d, d - 1) = 1$. Prove that there exists a unique invariant probability π, and the Markov process with this initial distribution is time-reversible. [*Hint*: There is a unique probability π for which (8.25) holds (with μ as counting measure). To solve for π, note that (8.25) implies $\pi(x + 1)/\pi(x) = \beta_x/\delta_{x+1}$ $(x = 0, 1, \ldots, d - 1)$. Check that this must be true of the ratios for any invariant probability.]

10. (*Time-Reversed Stationary Markov Process*) Let $\{X_n : n \geq 0\}$ be a stationary Markov process on $(S, \mathcal{S})$ with a transition probability density $p(x, y)$ (w.r.t. a σ-finite measure μ) and an invariant probability density $\pi(y)$, $y \in S$. Let $\{Z_n : -\infty < n < \infty\}$ be a stationary Markov process such that $\{Z_n : n \geq 0\}$ has the same distribution as $\{X_n : n \geq 0\}$. Show that $\{Y_n := Z_{-n} : n \in \mathbb{Z}\}$ is (i) stationary and (ii) Markov with the transition probability density $q(x, y) := (\pi(y)/\pi(x)) p(y, x)$; note that this is simply "Bayes formula" for (X_n, X_{n+1}). [*Hint*: Check that π is an invariant probability for q, and then construct a stationary double-sided Markov process $\{R_n\}_{-\infty < n < \infty}$ with these transition probabilities and invariant probability. Then check that the processes $\{Y_n : n \in \mathbb{Z}\}$ and $\{R_n : n \in \mathbb{Z}\}$ have the same distribution by considering finite dimensional events.]

11. Let $\{X_n : n \geq 0\}$ be a Markov chain on the countable state space S. Assume that for any $i, j \in S$, one has $p_{ij}^{(n)} > 0$ for some $n \geq 1$. We say that $p = ((p_{ij}))$ is *irreducible* in this case. Define $Y_n = (X_n, X_{n+1})$, $n = 0, 1, 2, \ldots$.

   (a) Show that $\{Y_n : n \geq 0\}$ is a Markov chain on $S' = \{(i, j) \in S \times S : p_{ij} > 0\}$.

   (b) Show that if $\{X_n : n \geq 0\}$ is irreducible, then so is $\{Y_n : n \geq 0\}$.

(c) Show that if $\{X_n : n \geq 0\}$ has invariant distribution $\boldsymbol{\pi} = (\pi_i)$, then $\{Y_n : n \geq 0\}$ has invariant distribution $(\pi_i\, p_{ij})$.

(d) Show that an irreducible Markov chain on a state space S with an invariant initial distribution $\boldsymbol{\pi}$ is time-reversible if and only if (*Kolmogorov Condition*):

$$p_{i i_1}\, p_{i_1 i_2} \cdots p_{i_k i} = p_{i i_k}\, p_{i_k i_{k-1}} \cdots p_{i_1 i} \qquad \text{for all } i, i_1, \ldots, i_k \in S, \quad k \geq 1.$$

(e) If there is a $j \in S$ such that $p_{ij} > 0$ for all $i \neq j$ in (d), then for time-reversibility it is both necessary and sufficient that $p_{ij}\, p_{jk}\, p_{ki} = p_{ik}\, p_{kj}\, p_{ji}$ for all i, j, k.

12. (*A General Finite State Space Graph*) Let $\{X_n : n \geq 0\}$ be an *irreducible* Markov chain on a finite state space S; i.e., for each $i, j \in S$, there is an $n \geq 1$ such that $p_{ij}^{(n)} > 0$. Define a *graph* G having states of S as vertices with edges joining i and j if and only if either $p_{ij} > 0$ or $p_{ji} > 0$.

(a) Show that G is connected; i.e., for any two sites i and j, there is a path of edges from i to j.

(b) Show that if $\{X_n : n \geq 0\}$ has an invariant distribution $\boldsymbol{\pi}$, then for any $A \subset S$,

$$\sum_{i \in A} \sum_{j \in S \setminus A} \pi_i\, p_{ij} = \sum_{i \in A} \sum_{j \in S \setminus A} \pi_j\, p_{ji}$$

i.e., the net probability flux across a *cut* of S into complementary subsets $A, S \setminus A$ is in balance. [*Hint*: Notice that $\sum_{i \in A} \sum_{j \in S} \pi_i\, p_{ij} = \sum_{i \in A} \sum_{j \in S} \pi_i\, p_{ji}$.]

(c) Show that if G contains no cycles of three or more vertices, i.e., $m = 3$ or more distinct vertices $v_1, \ldots, v_m$ such that v_i and v_{i+1} are joined by an edge for $i = 1, \ldots, m$ and $v_{m+1} = v_1$, then the process is time-reversible started with $\boldsymbol{\pi}$. A connected graph without cycles is called a *tree graph*. [*Hint*: Proceed inductively on the number of states.]

(d) Give a graphical proof that an invariant probability for a birth–death Markov chain on $\{0, 1, \ldots, N\}$ with reflecting boundaries at $0, N$ must be time-reversible.

13. Prove the time-reversibility of Example 3 when $|b| < 1$ and $\{\epsilon_n : n \geq 1\}$ is an i.i.d. standard normal sequence.

14. Consider a Markov process $\{X_n : n = 0, 1, 2, \ldots\}$ on a metric space S defined recursively by $X_{n+1} = g(X_n, \epsilon_{n+1}), n \geq 0$, where (i) $\{\epsilon_n : n \geq 1\}$ is a sequence of i.i.d. random variables with values in a metric space U, and independent of X_0, and (ii) $g : S \times U \rightarrow S$ is continuous. Show that the Markov process $\{X_n : n = 0, 1, 2, \ldots\}$ has the Feller property.