CAMBRIDGE
Cambridge University Press & Assessment
978-0-521-76840-5 — Ergodic Control of Diffusion Processes
Ari Arapostathis , Vivek S. Borkar , Mrinal K. Ghosh
Excerpt
More Information

1

Markov Processes and Ergodic Properties

1.1 Introduction

This book, as the title indicates, is about ergodic control of diffusion processes. The operative words here are *ergodic*, *control* and *diffusion processes*. We introduce two of these, diffusion processes and ergodic theory, in this chapter. It is a bird's eye view, sparse on detail and touching only the highlights that are relevant to this work. Further details can be found in many excellent texts and monographs available, some of which are listed in the bibliographical note at the end. The next level issue of control is broached in the next chapter.

We begin with diffusion processes, which is a special and important subclass of Markov processes. But before we introduce Markov processes, it is convenient to recall some of the framework of the general theory of stochastic processes which provides the backdrop for it.

Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a complete probability space, or in other words, Ω is a set called the sample space, $\mathfrak{F}$ is a σ-field of subsets of Ω (whence $(\Omega, \mathfrak{F})$ is a *measurable* space), and $\mathbb{P}$ is a probability measure on $(\Omega, \mathfrak{F})$. *Completeness* is a technicality that requires that any subset of a set having zero probability be included in $\mathfrak{F}$. A random process $\{X_t\}$ defined on $(\Omega, \mathfrak{F}, \mathbb{P})$ is a family of random variables indexed by a time index t which can be discrete (e.g., $t = 0, 1, 2, \dots$) or continuous (e.g., $t \geq 0$). We shall mostly be interested in the continuous time case. For notational ease, we denote by X the entire process. For each $t \in \mathbb{R}_+$, X_t takes values in a Polish space S, i.e., a separable Hausdorff space whose topology is metrizable with a complete metric. This is a convenient level of generality to operate with, because an amazingly large body of basic results in probability carry over to Polish spaces and most of the spaces one encounters in the study of random processes are indeed Polish. Let $d(\,\cdot\,,\,\cdot\,)$ be a complete metric on S. For any Polish space S, $\mathscr{P}(S)$ denotes the Polish space of probability measures on S under the Prohorov topology [32, chapter 2]. Recall that if S is a metric space, then a collection $M \subset \mathscr{P}(S)$ is called *tight* if for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset S$, such that $\mu(K_\varepsilon) \geq 1 - \varepsilon$ for all $\mu \in M$. The celebrated criterion for compactness in $\mathscr{P}(S)$, known as Prohorov's theorem, states that for a metric space S if $M \subset \mathscr{P}(S)$ is tight, then it is relatively compact and, provided S is complete and separable, the converse also holds [24, theorems 6.1 and 6.2, p. 37], [55, p. 104].

We shall usually have in the background a *filtration*, i.e., an increasing family $\{\mathfrak{F}_t\}$ of sub-σ-fields of $\mathfrak{F}$ indexed by t again. Intuitively, $\mathfrak{F}_t$ corresponds to *information available at time t*. Again, for technical reasons we assume that it is complete (i.e., contains all sets of zero probability and their subsets) and right-continuous (i.e., $\mathfrak{F}_t = \cap_{s>t}\mathfrak{F}_s$). We say that X is *adapted* to the filtration $\{\mathfrak{F}_t\}$ if for each t, X_t is $\mathfrak{F}_t$-measurable. A special case of a filtration is the so-called natural filtration of X, denoted by $\mathfrak{F}_t^X$ and defined as the completion of $\cap_{t'>t}\sigma(X_s : s \le t')$. Clearly X is adapted to its own natural filtration. An important notion related to a filtration is that of a *stopping time*. A $[0,\infty]$-valued random variable τ is said to be a stopping time w.r.t. the filtration $\{\mathfrak{F}_t\}$ (the filtration is usually implicit) if for all $t \ge 0$, $\{\tau \le t\} \subset \mathfrak{F}_t$. Intuitively, what this says is that at time t, one knows whether τ has occurred already or not. For example, the first time the process hits a prescribed closed set is a stopping time with respect to its natural filtration, but the last time it does so need not be. We associate with τ the σ-field $\mathfrak{F}_\tau$ defined by

$$\mathfrak{F}_\tau = \{A \in \mathfrak{F} : A \cap \{\tau \le t\} \in \mathfrak{F}_t \text{ for all } t \in [0,\infty)\}\ .$$

Intuitively, $\mathfrak{F}_\tau$ are the events prior to τ.

So far we have viewed X only as a collection of random variables indexed by t. But for a fixed sample point in Ω, it is also a function of t. The least we shall assume is that it is a measurable function. A stronger notion is progressive measurability, which requites that for each $T > 0$, the function $(t,\omega) \to X_t(\omega)$, $(t,\omega) \in [0,T]\times\Omega$, be measurable with respect to $\mathcal{B}_T \times \mathfrak{F}_T$, where $\mathcal{B}_T$ denotes the Borel σ-field on $[0,T]$. The sub-σ-field of $[0,\infty)\times\Omega$ generated by the progressively measurable processes is known as the progressively measurable σ-field. If a process is adapted to $\mathfrak{F}_t$ and has right or left-continuous paths, then it is progressively measurable [47, p. 89].

There is one serious technicality which has been glossed over here. Two random processes X, X' are said to be versions or modifications of each other if $X_t = X'_t$ a.s. for all t. This defines an equivalence relation and it is convenient to work with such equivalence classes. That is, when one says that X has measurable sample paths, it is implied that it has a version which is so. The stronger equivalence notion of $\mathbb{P}(X_t = X'_t \text{ for all } t \ge 0) = 1$ is not as useful. We shall not dwell on these technicalities too much. See Borkar [32, chapter 6], for details.

We briefly mention the issue of the actual construction of a random process. In practice, a random process is typically described by its finite dimensional marginals, i.e., the laws of $(X_{t_1},\dots,X_{t_n})$ for all finite collections of time instants $t_1 < \cdots < t_n$, $n \ge 1$. In particular, all versions of a process have the same finite dimensional distributions. These are perforce consistent, i.e., if $B \subset A$ are two such collections, then the law for B is the induced law from the law for A under the appropriate projection. The celebrated Kolmogorov extension theorem gives us the converse statement: Given a consistent family of such finite dimensional laws, there is a unique probability measure on $S^{[0,\infty)}$ consistent with it. Thus we can let $\Omega = S^{[0,\infty)}$, $\mathfrak{F}$ the product σ-field,

$\mathbb{P}$ denote this unique law, and set $X_t(\omega) = \omega(t)$, where $\omega = \{\omega(t) : t \geq 0\}$ denotes a sample point in Ω. This is called the canonical construction of the random process X. While it appears appealingly simple and elegant, it has its limitations. The most significant limitation is that $\mathfrak{F}$ contains only "countably described sets," i.e., any set in $\mathfrak{F}$ must be describable in terms of countably many time instants. (It is an interesting exercise to prove this.) This eliminates many sets of interest. Thus it becomes essential to "lift" this construction to a more convenient space, such as the space $C([0,\infty);S)$ of continuous functions $[0,\infty) \mapsto S$ or the space $\mathcal{D}([0,\infty);S)$ of functions $[0,\infty) \mapsto S$ that are continuous from the right and have limits from the left at each t (r.c.l.l.). We briefly sketch how to do the former, as that's what we shall need for diffusion processes.

The key result here is that if X is *stochastically continuous*, in other words if

$$\mathbb{P}\left(d(X_s, X_t) > \varepsilon\right) \to 0 \tag{1.1.1}$$

for any $\varepsilon > 0$ as $s \to t$, and for each $T > 0$, the modulus of continuity

$$w_T(X,\delta) := \sup\left\{d(X_s, X_t) : 0 \leq s \leq t \leq T,\ |t-s| < \delta\right\} \to 0 \quad \text{a.s.}, \tag{1.1.2}$$

as $\delta \to 0$, then X has a continuous version. The proof is simple: restrict X to rationals, extend it uniquely to a continuous function on $[0,\infty)$ (which is possible because (1.1.2) guarantees uniform continuity on rationals when restricted to any $[0,T]$, for $T > 0$), and argue using stochastic continuity that this indeed yields a version of X. A convenient test for (1.1.1) – (1.1.2) to hold is the Kolmogorov continuity criterion (see Wong and Hajek [122, pp. 57]), that for each $T > 0$ there exist positive scalars a, b, and c satisfying

$$\mathbb{E}\left[d(X_t, X_s)^a\right] \leq b|t-s|^{1+c} \qquad \forall\ t, s \in [0,T]\,.$$

Note that the above procedure a.s. defines a map that maps an element of Ω to the continuous extension of its restriction to the rationals. Defining the map to be the function that is identically zero on the zero probability subset of Ω that is left out, we have a measurable map $\Omega \mapsto C([0,\infty);S)$. The image μ of $\mathbb{P}$ under this map defines a probability measure on $C([0,\infty);S)$. We may thus realize the continuous version as a *canonically* defined random process X' on the new probability space $(C([0,\infty);S), \mathfrak{G}, \mu)$, where $\mathfrak{G}$ is the Borel σ-field of $C([0,\infty);S)$, as $X'_t(\omega) = \omega(t)$ for $\omega \in C([0,\infty);S)$. An analogous development is possible for the space $\mathcal{D}([0,\infty);S)$ of paths $[0,\infty) \to S$ that are right-continuous and have left limits. This space is Polish: it is separable and metrizable with a complete metric $\mathfrak{d}_s$ (due to Skorohod) defined as follows [55, p. 117]. Let Λ denote the space of strictly increasing Lipschitz continuous surjective maps λ from $\mathbb{R}_+$ to itself such that

$$\gamma(\lambda) := \operatorname*{ess\,sup}_{t\geq 0} |\log \lambda'(t)| < \infty\,.$$

With ρ a complete metric on S, define

$$d_T(x,y,\lambda) := \sup_{t\geq 0}\,\bigl[1 \wedge \rho(x(t\wedge T), y(\lambda(t)\wedge T))\bigr]\,, \qquad T>0\,,$$

$$\mathfrak{d}_s(x,y) := \inf_{\lambda\in\Lambda}\left[\gamma(\lambda) \vee \int_0^\infty \mathrm{e}^{-s} d_s(x,y,\lambda)\,\mathrm{d}s\right].$$

Convergence with respect to $\mathfrak{d}_s$ has the following simple interpretation: $x_n \to x$ in $\mathcal{D}([0,\infty); S)$ if there exists a sequence $\{\lambda_n \in \Lambda : n \in \mathbb{N}\}$ such that $\lambda_n(t) \to t$ uniformly on compacta, and $\sup_{[0,T]}\ \rho(x_n \circ \lambda_n, x) \to 0$ for all $T > 0$. This topology is known as the Skorohod topology.

A useful criterion due to Chentsov for the existence of a r.c.l.l. version that extends the Kolmogorov continuity criterion is that for any $T > 0$ there exist positive constants a, b, c and C satisfying [66, pp. 159–164]

$$\mathbb{E}\Bigl[|X_t - X_r|^a |X_r - X_s|^b\Bigr] \leq C|t-s|^{1+c}\,, \qquad \forall s<r<t\,.$$

1.2 Markov processes

Before defining Markov processes, it is instructive to step back and recall what a deterministic dynamical system is. A deterministic dynamical system has as its backdrop a set Σ called the *state space* in which it evolves. Its evolution is given by a *time t map* Φ_t, $t \in \mathbb{R}$, with the interpretation that for $x \in \Sigma$, $x(t) := \Phi_t(x)$ is the position of the system at time t if it starts at x at time 0. The idea is that once at x, the trajectory $\{x(t) : t \geq 0\}$ is completely specified, likewise for $t \leq 0$. This in fact is what qualifies $x(t)$ as the *state at time t* in the sense of physics: $x(t)$ is all you need to know at time t to be able to determine the future trajectory $\{x(s) : s \geq t\}$. Thus $\Phi_0(x) = x$, and $\Phi_t \circ \Phi_s = \Phi_s \circ \Phi_t = \Phi_{s+t}$, i.e., $\{\Phi_t : t \in \mathbb{R}\}$ is a group.

A two-parameter version is possible for time-dependent dynamics, i.e., when the future (or past) trajectory depends on the position x as well as the precise time t_0 at which the trajectory is at x. Thus we need a two-parameter group $\Phi_{s,t}$, satisfying $\Phi_{t,t}(x) = x$, and $\Phi_{s,t} \circ \Phi_{u,s} = \Phi_{u,t}$ for all u, s, and t.

Clearly for stochastic dynamical systems, it does not make sense to demand that the complete future trajectory be determined by the present position. Nevertheless there is a natural generalization of the notions of a state and a dynamical system. We require that the (regular) conditional law of $\{X_s : s \geq t\}$, given $\{X_u : u \leq t\}$, should be the same as its conditional law given X_t alone. In other words, knowing how one arrived at X_t tells us nothing more about the future than what is already known by knowing X_t. From an equivalent definition of conditional independence [32, p. 42], this is equivalent to the statement that $\{X_s : s > t\}$ and $\{X_s : s < t\}$ are conditionally independent given X_t for each t. This definition is symmetric in time. Thus, for example, it is also equivalent to: for each t, the regular conditional law of $\{X_s : s < t\}$, given $\{X_s : s \geq t\}$, is the same as its regular conditional law given X_t. In fact, a more general statement is true: for any $t_1 < t_2$, $\{X_s : s < t_1 \text{ or } s > t_2\}$ and

$\{X(s) : t_1 < s < t_2\}$ are conditionally independent given $\{X_{t_1}, X_{t_2}\}$. This also serves as an equivalent definition. Though not very useful in the present context, this is the definition that extends well to indices more general than time, such as $t \in \mathbb{R}^2$. These are the so-called Markov random fields.

Since finite dimensional marginals completely specify the law of a stochastic process, an economical statement of the Markov property is: For every collection of times $0 \le t \le t_1 < \cdots < t_k < \infty$, and Borel subsets $A_1, \dots, A_k \subset S$, it holds that

$$\mathbb{P}\left(X_{t_i} \in A_i,\ i = 1, \dots, k \mid \mathfrak{F}^X_t\right) = \mathbb{P}\left(X_{t_i} \in A_i,\ i = 1, \dots, k \mid X_t\right) .$$

A stronger notion is that of the *strong Markov property*, which requires that

$$\mathbb{P}\left(X_{\tau+t_i} \in A_i,\ i = 1, \dots, k \mid \mathfrak{F}^X_\tau\right) = \mathbb{P}\left(X_{\tau+t_i} \in A_i,\ i = 1, \dots, k \mid X_\tau\right)$$

a.s. on $\{\tau < \infty\}$, for every $(\mathfrak{F}^X_t)$-stopping time τ. If X has the Markov, or strong Markov property, then it said to be a Markov, or a strong Markov process, respectively.

For $t > s$ and $x \in S$, let $P(s, x, t, \mathrm{d}y)$ denote the regular conditional law of X_t given $X_s = x$. This is called the transition probability (kernel) of X. In particular, $P\colon (s, x, t) \mapsto P(s, x, t, S)$ is measurable. By the filtering property of conditional expectations:

$$\mathbb{E}\left[\mathbb{E}\left[f(X_t) \mid \mathfrak{F}^X_r\right] \bigm| \mathfrak{F}^X_s\right] = \mathbb{E}\left[f(X_t) \mid \mathfrak{F}^X_s\right] , \qquad s \le r \le t ,$$

which, coupled with the Markov property, yields

$$P(s, x, t, \mathrm{d}y) = \int_S P(s, x, r, \mathrm{d}z)\, P(r, z, t, \mathrm{d}y) , \qquad s \le r \le t .$$

These are called the Chapman–Kolmogorov equations. While the transition probability kernels of Markov processes must satisfy these, the converse is not true [57].

Let $\mathcal{B}(S)$ denote the space of bounded measurable functions $S \mapsto \mathbb{R}$. Define

$$T_{s,t} f(x) := \int P(s, x, t, \mathrm{d}y)\, f(y) , \qquad t \ge s \ge 0 , \qquad f \in \mathcal{B}(S) .$$

Then by the Chapman–Kolmogorov equations, $\{T_{s,t} : 0 \le s \le t\}$ is a two-parameter semigroup of operators, i.e., it satisfies

$$T_{t,t} = I , \qquad T_{r,t} \circ T_{s,r} = T_{s,t} , \qquad 0 \le s \le r \le t ,$$

where I is the identity operator. This is weaker than the group property for deterministic flows. However, this is inevitable because of the irreversibility of stochastic processes.

Let $C_b(S)$ denote the space of bounded continuous real-valued functions on S. The process X above is said to be *Feller* if $\{T_{s,t} : 0 \le s \le t\}$ maps $C_b(S)$ into $C_b(S)$, and *strong Feller* if it maps $\mathcal{B}(S)$ into $C_b(S)$. The former case is obtained if the transition probability kernel $P(s, x, t, \mathrm{d}y)$ is continuous in the initial state x. The latter requires more – the kernel should have some additional smoothing properties.

For example, when $S = \mathbb{R}$, the Gaussian kernel

$$P(s, x, t, \mathrm{d}y) = \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-\frac{(x-y)^2}{2(t-s)}} \mathrm{d}y$$

does have this property. More generally, if the transition kernel is of the form

$$P(s, x, t, \mathrm{d}y) = p(s, x, t, y)\lambda(\mathrm{d}y)$$

for a positive measure λ and the density p is continuous in the x variable, then the strong Feller property follows by the dominated convergence theorem.

An important consequence of the Feller property is that it implies the strong Markov property for Markov processes with right-continuous paths. To see this, one first verifies the strong Markov property for stopping times taking values in $\left\{\frac{k}{2^n} : k \geq 0\right\}$ for $n \in \mathbb{N}$ fixed. This follows by a straightforward verification [55, p. 159] which does not require the Feller property. Given a general a.s.-finite stopping time τ, the property then holds for $\tau^{(n)} := \frac{\lfloor 2^n \tau \rfloor + 1}{2^n}$ which is also seen to be a stopping time for all $n \geq 1$. As $n \uparrow \infty$, $\tau^{(n)} \downarrow \tau$. Using the right-continuity of paths along with the Feller property and the a.s. convergence property of reverse martingales, the strong Markov property for τ can be inferred from that for $\tau^{(n)}$.

In the case of Feller processes, we may restrict $\{T_{s,t}\}$ to a semigroup on $C_b(S)$. Of special interest is the case when the transition probability kernel $P(s, x, t, \mathrm{d}y)$ depends only on the difference $r = t - s$. By abuse of terminology, we then write $P(r, x, \mathrm{d}y)$ and also $T_r = T_{s,s+r}$. Then T_t, $t \geq 0$ defines a one-parameter semigroup of operators. A very rich theory of such semigroups is available, with which we deal in the next section. It is worth noting here that this is a special case of the general theory of operator semigroups. These are the so-called Markov semigroups, characterized by the additional properties:

(a) $T_t(\alpha f + \beta g) = \alpha T_t f + \beta T_t g$ for all $\alpha, \beta \in \mathbb{R}$ and $f, g \in C_b(S)$;
(b) $f \geq 0 \implies T_t f \geq 0$;
(c) $\|T_t f\| \leq \|f\|$, where $\|f\| := \sup_{s \in S} |f(s)|$;
(d) $T_t \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the constant function $\equiv 1$.

We now give some important examples of Markov processes.

Example 1.2.1 (i) Poisson process: Let $\lambda > 0$. A $\mathbb{Z}_+$-valued stochastic process $N = \{N_t : t \geq 0\}$ is called a Poisson process with parameter λ if

(a) $N_0 = 0$;
(b) for any $0 \leq t_0 < t_1 < \cdots < t_n$,

$$N_{t_1} - N_{t_0},\ N_{t_2} - N_{t_1}, \ldots,\ N_{t_n} - N_{t_{n-1}}$$

are independent, i.e., N has independent increments;
(c) $t \mapsto N_t$ is a.s. right-continuous;

(d) for $t \geq s \geq 0$, $N_t - N_s$ has the Poisson distribution with parameter λ, i.e.,

$$\mathbb{P}\,(N_t - N_s = n) = \frac{\lambda^n (t-s)^n \mathrm{e}^{-\lambda(t-s)}}{n!}\,, \qquad n = 0, 1, \dots$$

It can be verified that N is a Markov process with transition function

$$P(t, m, \{n\}) = \frac{(\lambda t)^{n-m} \mathrm{e}^{-\lambda t}}{(n-m)!}\,, \qquad n \geq m\,.$$

(ii) One-dimensional Brownian motion: A real-valued process $\mathbb{B} = \{\mathbb{B}_t : t \geq 0\}$ is called a one-dimensional Brownian motion (or Wiener process) if

(a) $\mathbb{B}_0 = 0$;
(b) $t \mapsto \mathbb{B}_t$ is a.s. continuous;
(c) for $t > s \geq 0$, $\mathbb{B}_t - \mathbb{B}_s$ has the Normal distribution with mean 0 and variance $t - s$;
(d) for any $0 \leq t_0 < t_1 < \cdots < t_n$,

$$\mathbb{B}_{t_1} - \mathbb{B}_{t_0},\ \mathbb{B}_{t_2} - \mathbb{B}_{t_1}, \dots,\ \mathbb{B}_{t_n} - \mathbb{B}_{t_{n-1}}$$

are independent.

Then $\mathbb{B}$ is a Markov process with transition function

$$P(t, x, A) = \frac{1}{\sqrt{2\pi t}} \int_A \mathrm{e}^{-\frac{(y-x)^2}{2t}}\, \mathrm{d}y\,.$$

(iii) In general, any stochastic process with independent increments is a Markov process.

(iv) d-dimensional Brownian motion: An $\mathbb{R}^d$-valued process $W = \{W_t : t \geq 0\}$, where $W_t = (W_t^1, \dots, W_t^d)$, is called a d-dimensional Brownian motion if

(a) for each i, $W^i = \{W_t^i : t \geq 0\}$ is a one-dimensional Brownian motion;
(b) for $i \neq j$, the processes W^i and W^j are independent.

Then W is a Markov process with transition function

$$P(t, x, A) = \frac{1}{(2\pi t)^{d/2}} \int_A \mathrm{e}^{-\frac{|y-x|^2}{2t}}\, \mathrm{d}y\,.$$

There are several ways of constructing Markov processes. We list the common ones below.

(1) *Via the theorem of Ionescu–Tulcea:* Once an initial law λ at t_0 and the transition probability kernel are prescribed, one can write down the finite dimensional marginals of the process:

$$\mathbb{P}\,(X_{t_k} \in A_k,\ 0 \leq k \leq n) = \int_{A_0} \lambda(\mathrm{d}y_0) \int_{A_1} P(t_0, y_0, t_1, \mathrm{d}y_1) \cdots \int_{A_n} P(t_{n-1}, y_{n-1}, t_n, \mathrm{d}y_n)\,.$$

These are easily seen to form a consistent family and thus define a unique law for X, by the theorem of Ionescu–Tulcea [93, p. 162]. This may be lifted to a suitable function space such as $C([0,\infty); S)$ by techniques described in Section 1.1.

(2) *Via dynamics driven by an independent increment process:* A typical instance of this is the equation

$$X_t = X_0 + \int_0^t h(X_s)\,\mathrm{d}s + W_t\,, \qquad t \ge 0\,,$$

where W is a Brownian motion. Suppose that for a given trajectory of W, this equation has an a.s. unique solution X (this requires suitable hypotheses on h). Then for $t > s$,

$$X_t = X_s + \int_s^t h(X_r)\,\mathrm{d}r + W_t - W_s\,, \qquad t \ge 0\,,$$

and therefore X_t is a.s. specified as a functional of X_s and the independent increments $\{W_u - W_s : s \le u \le t\}$. The Markov property follows easily from this. Later on we shall see that the a.s. uniqueness property used above holds for a very general class of equations. One can also consider situations where W is replaced by other independent increment processes.

(3) *Via change of measure:* Suppose X is a Markov process constructed canonically on its path space, say $C([0,\infty); S)$. That is, we take $\Omega = C([0,\infty); S)$, $\mathfrak{F}$ its Borel σ-field, and $\mathbb{P}$ the law of X. Then $X_t(\omega) = \omega(t)$ for $t \ge 0$ and $\omega \in \Omega$. Let $X_{[s,t]}$ denote the trajectory segment $\{X_r : s \le r \le t\}$. Let $\mathfrak{F}^X_{s,t}$ be the right-continuous completion of $\sigma\{X_r : s \le r \le t\}$. A family $\{\Lambda_{s,t} : s < t\}$ of $\mathfrak{F}^X_{s,t}$-measurable random variables is said to be a multiplicative functional if $\Lambda_{r,s}\Lambda_{s,t} = \Lambda_{r,t}$ for all $r < s < t$. If in addition $\{\Lambda_{0,t} : t \ge 0\}$ is a nonnegative martingale with mean equal to one, we can define a new probability measure $\hat{\mathbb{P}}$ on $(\Omega, \mathfrak{F})$ as follows: If $\mathbb{P}_t$ and $\hat{\mathbb{P}}_t$ denote the restrictions of $\mathbb{P}$ and $\hat{\mathbb{P}}$, respectively, to $\mathfrak{F}^X_t$ for $t \ge 0$, then

$$\frac{\mathrm{d}\hat{\mathbb{P}}_t}{\mathrm{d}\mathbb{P}_t} = \Lambda_{0,t}\,, \qquad t \ge 0\,. \tag{1.2.1}$$

Since $\mathfrak{F} = \bigvee_{t\ge 0} \mathfrak{F}^X_t$, it follows by the martingale property of $\{\Lambda_{0,t} : t \ge 0\}$ that (1.2.1) consistently defines a probability measure on $(\Omega, \mathfrak{F})$. Let $\mathbb{E}$ and $\hat{\mathbb{E}}$ denote the expectations under $\mathbb{P}$ and $\hat{\mathbb{P}}$, respectively. For any $f \in \mathcal{B}(S)$ and $s < t$, one has the well-known *Bayes formula*

$$\hat{\mathbb{E}}\left[f(X_t) \mid \mathfrak{F}^X_s\right] = \frac{\mathbb{E}\left[f(X_t)\Lambda_{0,t} \mid \mathfrak{F}^X_s\right]}{\mathbb{E}\left[\Lambda_{0,t} \mid \mathfrak{F}^X_s\right]} = \frac{\mathbb{E}\left[f(X_t)\Lambda_{0,t} \mid \mathfrak{F}^X_s\right]}{\Lambda_{0,s}}\,.$$

From the multiplicative property, the right-hand side is simply

$$\mathbb{E}\left[f(X_t)\Lambda_{s,t} \mid \mathfrak{F}^X_s\right] = \mathbb{E}\left[f(X_t)\Lambda_{s,t} \mid X_s\right]\,,$$

in other words, a function of X_s alone. Thus X remains a Markov process under $\hat{\mathbb{P}}$. We shall later see an important instance of this construction when we construct the so-called weak solutions to stochastic differential equations.

(4) *Via approximation:* Many Markov processes are obtained as limits of simpler Markov (or even non-Markov) processes. In the next section, we discuss the semigroup and martingale approaches to Markov processes. These provide the two most common approximation arguments in use. In the semigroup approach, one works with the semigroup $\{T_t\}$ described above. This specifies the transition probability kernel via

$$\int P(t, x, \mathrm{d}y) f(y) = T_t f(x)\,, \qquad f \in C_b(S)\,,$$

and therefore also determines the law of X once the initial distribution is given. One often constructs Markov semigroups $\{T_t\}$ as limits (in an appropriate sense) of a sequence of known Markov semigroups $\{T_t^n : n \geq 1\}$ as $n \uparrow \infty$. See Ethier and Kurtz [55] for details and examples. The martingale approach, on the other hand, uses the *martingale characterization*, which characterizes the Markov process X by the property that for a sufficiently rich class of $f \in C_b(S)$ and an operator $\mathcal{L}$ defined on this class, $f(X_t) - \int_0^t \mathcal{L}f(X_s)\,\mathrm{d}s$, $t \geq 0$, is an $(\mathfrak{F}_t^X)$-martingale. As the martingale property is preserved under weak (Prohorov) convergence of probability measures, this allows us to construct Markov processes as limits in law of other Markov or sometimes non-Markov processes. The celebrated *diffusion limit* in queuing theory is a well-known example of this scheme, as are many systems of infinite interacting particles.

We conclude this section by introducing the notion of a *Markov family*. This is a family of probability measures $\{\mathbb{P}_x : x \in S\}$ on $(\Omega, \mathfrak{F})$ along with a stochastic process X defined on it such that the law of X under $\mathbb{P}_x$ for each x is that of a Markov process corresponding to a common transition probability kernel (i.e., with a common functional dependence on x), with initial condition $X_0 = x$. This allows us to study the Markov process under multiple initial conditions at the same time. We denote a Markov family by $(X, (\Omega, \mathfrak{F}), \{\mathbb{P}_x\}_{x\in S})$.

1.3 Semigroups associated with Markov processes

Let E be a Polish space, and $(X, (\Omega, \mathfrak{F}), \{\mathfrak{F}_t\}_{t\in\mathbb{R}_+}, \{\mathbb{P}_x\}_{x\in E})$ a Markov family. We define a one-parameter family of operators $T_t \colon \mathcal{B}(E) \to \mathcal{B}(E)$, $t \in \mathbb{R}_+$, as follows:

$$T_t f(x) := \mathbb{E}_x\left[f(X_t)\right] = \int_E P(t, x, \mathrm{d}y) f(y)\,, \qquad f \in \mathcal{B}(E)\,. \tag{1.3.1}$$

The following properties are evident

(i) for each t, T_t is a linear operator;
(ii) $\|T_t f\|_\infty \leq \|f\|_\infty$, where $\|\cdot\|_\infty$ is the L^∞-norm;
(iii) T_t is a positive operator, i.e., $T_t f \geq 0$ if $f \geq 0$;
(iv) $T_t\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the function identically equal to 1;
(v) $T_0 = I$, where I denotes the identity operator;
(vi) $T_{t+s} = T_t T_s$.

Thus $\{T_t : t \geq 0\}$ is a contractive semigroup of positive operators on $\mathcal{B}(E)$. We next define another semigroup, which is in a sense dual to $\{T_t : t \geq 0\}$.

Let $\mathfrak{M}_s(E)$ denote the space of all finite signed measures on $(E, \mathscr{B}(E))$, under the topology of total variation norm. For $t \in \mathbb{R}_+$, we define $S_t \colon \mathfrak{M}_s(E) \to \mathfrak{M}_s(E)$ by

$$(S_t\mu)(A) := \int_E P(t, x, A)\,\mu(\mathrm{d}x)\,.$$

Then properties (i)–(vi) hold for $\{S_t\}$, under the modification $\|S_t\mu\|_{\mathrm{TV}} \leq \|\mu\|_{\mathrm{TV}}$ for property (ii) and $S_t\mu(E) = \mu(E)$ for property (iv).

Let $f \in \mathcal{B}(E)$. We define an operator $\mathcal{A}$ on $\mathcal{B}(E)$ by

$$\mathcal{A}f = \lim_{t \downarrow 0} \frac{T_t f - f}{t}$$

in L^∞-norm, provided the limit exists. We refer to the set of all such functions f as the domain of $\mathcal{A}$ and denote it by $\mathscr{D}(\mathcal{A})$. The operator $\mathcal{A}$ is called the *infinitesimal generator* of the semigroup $\{T_t\}$. Let

$$\mathcal{B}_0 := \{f \in \mathcal{B}(E) : \|T_t f - f\| \to 0, \text{ as } t \downarrow 0\}\,.$$

It is easy to verify that $\mathcal{B}_0$ is a closed subspace of $\mathcal{B}(E)$ and that $T_t f$ is uniformly continuous in t for each $f \in \mathcal{B}_0$. Also $T_t(\mathcal{B}_0) \subset \mathcal{B}_0$ for all $t \in \mathbb{R}_+$ and $\mathscr{D}(\mathcal{A}) \subset \mathcal{B}_0$. Thus $\{T_t\}$ is a strongly continuous semigroup on $\mathcal{B}_0$. For Feller processes we have $T_t \colon C_b(E) \to C_b(E)$ and the previous discussion applies with $C_b(E)$ replacing $\mathcal{B}(E)$.

The following result is standard in semigroup theory.

Proposition 1.3.1 *For a strongly continuous semigroup $\{T_t\}$ on a Banach space $\mathcal{X}$ with generator $\mathcal{A}$, the following properties hold:*

(i) *if $f \in \mathcal{X}$, then $\int_0^t T_s f\,\mathrm{d}s \in \mathscr{D}(\mathcal{A})$ for all $t \in \mathbb{R}_+$, and*

$$T_t f - f = \mathcal{A} \int_0^t T_s f\,\mathrm{d}s\,;$$

(ii) *if $f \in \mathscr{D}(\mathcal{A})$, then $T_t f \in \mathscr{D}(\mathcal{A})$ for all $t \in \mathbb{R}_+$, and*

$$\frac{\mathrm{d}}{\mathrm{d}t} T_t f = \mathcal{A} T_t f = T_t \mathcal{A} f\,,$$

or equivalently,

$$T_t f - f = \int_0^t \mathcal{A} T_s f\,\mathrm{d}s = \int_0^t T_s \mathcal{A} f\,\mathrm{d}s\,, \qquad t \in \mathbb{R}_+\,.$$

We describe briefly the construction of a semigroup from its infinitesimal generator $\mathcal{A}$. For this purpose we introduce the notion of the *resolvent* of a semigroup $\{T_t\}$. This is a family of operators $\{\mathcal{R}_\lambda\}_{\lambda > 0}$ on $\mathcal{B}(E)$ defined by

$$\mathcal{R}_\lambda f(x) = \int_0^\infty \mathrm{e}^{-\lambda t} T_t f(x)\,\mathrm{d}t = \mathbb{E}_x\left[\int_0^\infty \mathrm{e}^{-\lambda t} f(X_t)\,\mathrm{d}t\right], \qquad x \in E\,.$$