Markov Processes, Contraction Semigroups, and Infinitesimal Generators

Alex Kazachek

Abstract

Markov processes are stochastic processes with restrictions on their conditional probabilities, so that their updates depend only on their most recent realizations. Each Markov process is intimately connected with an operator known as its infinitesimal generator, allowing one to recover the process (or rather, its semigroup) through exponentiation. We show the two-way relationship between these processes and their generators, the content of the celebrated Hille-Yosida and Lumer-Phillips theorems. This involves a focus on Feller processes, of which the canonical Markov process – the Weiner process – is an instance, and we explore its infinitesimal generator in the one-dimensional case.

22 March 2024

1 Introduction

Markov processes are indispensable in the theory and application of stochastic processes. Information theory studies the ability of parties to communicate via some channel; the original formulation by Shannon [Sha48], and the still-standing modern definition, is that a channel is described by a Markov transition matrix. Reinforcement learning is a machine learning paradigm wherein an agent attempts to learn a task through repeated attempts with a positive reward for success; one of its most successful formalisms is of the agent partaking in a Markov decision process which may be optimized through Q-learning [MRT18]. As physical systems are almost always Markovian, changing and reacting based only on their present state, Markov processes are used in modelling various physical phenomenon such as chemical kinetics [Sco13]. The Weiner process, the continuous-time limit of a discrete random walk, is a Markov process which forms the backbone of mathematical finance as the starting model for an asset price is that of geometric Brownian motion [Shr04]. This notion has even made its way to political ecology, in proposals of structures to replace the hegemonic delineation of humanity and nature [Lat04].

Little more needs to be said of the vitality Markov processes imbue in the mathematical and physical sciences. Indeed, perhaps not even the above was necessary; Markovian dynamics are ubiquitous. And due to their ubiquity the mathematical theory and formalism of Markov processes has far exceeded what Markov could have imagined when first applying them to study vowel-consonant pairs in poetry [Hay13]. We will build up Markov processes over metric spaces, and then naturally derive the notion of an infinitesimal generator, allowing us to describe the process by a differential equation. Along the way, we will show this is not merely a description, however, but a characterization! Sufficiently nice infinitesimal generators yield quite nice Markov processes, known as Feller processes. These encapsulate a large number of common Markov processes, such as the Weiner process, which we will use as our concluding example, showing that the diffusion equation is its generator.

2 Mathematical Background

We will begin by recalling some facts from functional analysis and measure theory, and establishing our notation. Any standard reference will do here, such as [SR80; Bre10; Bas22].

2.1 Measuring and Integrating

Denote throughout by (E, d) a metric space. Its standard topology is generated by its open balls. We will assume that E is locally compact.

Definition 1 (Measurable). *Let $B(E)$ be the Borel σ-algebra over E. Denote by $\mathscr{B}(E)$ the set of all measurable functions. That is,*

$$\mathscr{B}(E) = \{f : E \to \mathbb{R} : f^{-1}(V) \in B(E) \text{ for all open } V\}.$$

We will focus, eventually, on the following functions. They have a nice duality which lets us associate them with measures, and they optimize nicely.

Definition 2 (Vanishing at Infinity). *Let $\mathscr{C}(E)$ be all continuous functions $E \to \mathbb{R}$. Further define $\mathscr{C}_0(E) \subseteq \mathscr{C}(E)$ as the set of all functions vanishing at infinity, meaning there is some $x_0 \in E$ such that for any $\varepsilon > 0$ we can find some R so $|f(x)| > \varepsilon$ whenever $d(x, x_0) > R$.*

We declare that a generic Banach space will be $(X, \|\cdot\|)$, with $\mathbb{C}$ the underlying field and identity $\mathbb{1}$. Note that $\mathscr{B}(E)$, $\mathscr{C}(E)$, and $\mathscr{C}_0(E)$ are all Banach when endowed with the infinity-norm

$$\|f\|_\infty = \sup_{x \in E} |f(x)|,$$

and we will take this to be the implicit norm for all these spaces.

Lemma 1. *Let $f \in \mathscr{C}_0(E)$. Then, there exists some $x \in E$ such that $|f(x)| = \|f\|$.*

Proof. Suppose $\|f\| > 0$, otherwise any $x \in E$ works. By definition of vanishing at infinity, we can find some $x_0 \in E$ and $R > 0$ such that $|f(x)| \leq \|f\|/2$ whenever $d(x, x_0) \geq R$. Therefore,

$$K = \{x \in E : \|f\|/2 \leq x \leq \|f\|\} \subseteq \{x \in E : d(x, x_0) \leq R\}.$$

In particular, we find that K is compact. Due to continuity of f, we know it attains its maximum on K. □

Theorem 1 (Riesz-Markov-Kakatuni). *The dual of $\mathscr{C}_0(E)$ is all finite (real!) measures on $B(E)$. Specifically, for such a measure μ we have the form*

$$\langle f, \mu \rangle = \int_E f(x) \, \mathrm{d}\mu(x).$$

We will have to integrate many things. The standard tool for this is the Lebesgue integral, though we will have to slightly generalize this to the Bochner integral. A generic measure space henceforth will be denoted by the triple (S, Σ, μ) (and just (S, Σ) will be our measurable space). When $S = \mathbb{R}$, we will write $\mathrm{d}x = \mathrm{d}\mu$ for the Lebesgue measure.

Definition 3 (Measurable). *Let $f : (S, \Sigma) \to (S', \Sigma')$ be a map between two measurable spaces. We say f is Σ-measurable if $f^{-1}(A) \in \Sigma$ for all $A \in \Sigma'$.*

Definition 4 (L^p)**.** *Let $f: S \to \mathbb{R}$ be a function measurable on the space (S, Σ), the set of all such functions being $M(S, \Sigma)$. For $1 \le p < \infty$ we define*

$$L^p(S, \Sigma, \mu) = \left\{ f \in M(S, \Sigma) : x \int_S |f|^p \, \mathrm{d}\mu < \infty \right\}.$$

For $p = \infty$, we define

$$L^\infty(S, \Sigma, \mu) = \left\{ f \in M(S, \Sigma) : \operatorname{ess\,sup}_{x \in S} f(x) < \infty \right\}.$$

Above, we of course have the Lebesgue integral. However, at some later point we will have to make use of a slightly more general one – the Bochner integral. This is defined in essentially the same as the Lebesgue integral, except that our functions may take on values in a Banach space.

Definition 5 (Bochner Integral)**.** *For $i = 1, \ldots, n$ let $A_i \in \Sigma$ be pairwise-disjoint sets of finite measure and take some $x_i \in X$. Define $f: S \to X$ by*

$$f(x) = \sum_{i=1}^{n} \chi_{A_i}(x) x_i,$$

a simple function. Here, χ_{A_i} are characteristic functions of A_i. Its integral is

$$\int_S f \, \mathrm{d}\mu = \sum_{i=1}^{n} \mu(A_i) x_i.$$

For a generic $f: S \to X$, if there exists some sequence of simple s_n such that

$$\int_S \|f - s_n\| \, \mathrm{d}\mu \to 0$$

then f is integrable and its integral is

$$\int_S f \, \mathrm{d}\mu = \lim_{n \to \infty} \int_S s_n \, \mathrm{d}\mu.$$

This Bochner integral has no differences to the Lebesgue integral which will be material to us. In fact, we may even define the L^p spaces above for functions $f: S \to X$ as well, provided we are slightly more careful about what it means for a function to be measurable with respect to a Banach space. In any case, the following theorem does apply with no changes [Hun14].

Theorem 2 (Lebesgue Differentiation)**.** *Let $f: \mathbb{R} \to X$ be integrable. Then,*

$$f(t) = \lim_{h \to 0} \frac{1}{h} \int_t^{t+h} f(x) \, \mathrm{d}x$$

holds for almost all t.

2.2 Operators and Spectra

We will denote by $\mathcal{L}(X)$ bounded linear operators $X \to X$ (for we will never need the domain and codomain to vary). On $\mathcal{L}(X)$ we have the operator norm which makes it Banach:

$$\|T\|_{\mathrm{op}} = \sup_{0 \ne x \in X} \frac{\|Tx\|}{\|x\|}.$$

The etymology of a bounded operator is precisely in that the operator norms are finite. Often, though, we define some norm and then restrict our interests to a set for which that norm is finite, excluding elements with infinite norms. We may do this, in some sense, for operators as well. The non-trivial hurdle is that if an operator has infinite norm, there is some point where it is discontinuous, and therefore cannot be linear, and so cannot have been an operator in the first place. We resolve this with the following definition.

Definition 6 (Unbounded Operator)**.** *Let $Y \subseteq X$ be a dense linear subspace. Then, a linear operator $T\colon Y \to X$ is a called an unbounded operator. We often write $Y = \operatorname{dom} T$. We say T is closed if its graph*

$$\Gamma(T) = \{(x, Tx) : x \in \operatorname{dom} T\}$$

is closed.

Of course, we could take $\operatorname{dom} T = X$ and then have an ordinary bounded operator present itself as a special case. Working with unbounded operators is slightly tricky, For instance, it is unclear in general if squaring makes sense since the operator may map outside of its own domain.

Definition 7 (Spectrum)**.** *Let $T \in \mathcal{L}(X)$. We define its spectrum*

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda \mathbb{1} \notin \mathcal{L}(X)\}.$$

There are multiple reasons $T - \lambda \mathbb{1}$ may fail to be a linear operator. The simplest one is when an eigenvalue exists, in which case the set-theoretic inverse simply does not exist. We may be tempted to define $\sigma(T)$ for an unbounded T in the same way. However, inversion is more subtle since that would be a map $X \to \operatorname{dom} T$. Instead, we use the following approach.

Definition 8 (Resolvent)**.** *Let T be an unbounded operator on X. Define its resolvent set*

$$\rho(T) = \{\lambda \in \mathbb{C} : R_\lambda(T) = (T - \lambda \mathbb{1})^{-1} \in \mathcal{L}(X)\}.$$

We define the spectrum of T by $\sigma(T) = \mathbb{C} \setminus \rho(T)$.

3 Markov Processes

We follow the presentation of [Gui20], filling in gaps (and correcting a few typos!) in our study of Markov and Feller processes. Just prior we have a brief review of probability theory, with the suggested reference being [Dur19].

3.1 Probability Primer

A probability measure is simply a measure space $(\Omega, \Sigma, \mathbb{P})$ where $\mathbb{P}(\Omega) = 1$. We will not restrict ourselves to random variables solely taking on real values, but will be working with metric spaces throughout.

Definition 9 (Random Variable)**.** *A random variable $Y\colon \Omega \to E$ is an Ω-measurable function. For Borel $A \in B(E)$ we use the notation*

$$\mathbb{P}(Y \in A) = \mathbb{P}(Y^{-1}(A))$$

and, if it exists, denote the expectation by

$$\mathbb{E}(Y) = \int_\Omega Y \, \mathrm{d}\mathbb{P}.$$

Definition 10 (Conditional Expectation)**.** *Let $\tilde{\Sigma} \subseteq \Sigma$ be a sub-σ-algebra. We define the conditional expectation of Y given $\tilde{\Sigma}$ as the unique $\tilde{\Sigma}$-measurable random variable $\mathbb{E}(Y \mid \tilde{\Sigma})$ such that*

$$\int_A Y \, \mathrm{d}\mathbb{P} = \int_A \mathbb{E}(Y \mid \tilde{\Sigma}) \, \mathrm{d}\mathbb{P}$$

for all $A \in \tilde{\Sigma}$.

Of course, the existence and uniqueness of such a random variable is actually a theorem, and not a definition. Throughout our study of Markov processes we will be conditioning on various sub-σ-algebras (with the fancy way to say this being adapting to filtrations), and so will make extensive use of this definition.

3.2 Markov Processes

A key space in the study of Markov processes is that of trajectories. The approach may seem slightly backward at first, but here is the idea: we can begin with some deterministic path, eating in time and regurgitating a location, and then at each point in our space, look at the probability that some path has led us here.

Definition 11 (Skorokhod Space)**.** *We say a function $\gamma : [0, \infty) \to E$ is càdlàg (continue à droite, limite à gauche) if it is right-continuous with left limits. We then define the Skorohod space*

$$D_E[0, \infty) = \{\gamma : [0, \infty) \to E : \gamma \text{ is càdlàg}\},$$

and call these γ trajectories.

We also define the time-mapping functions $\pi_t(\gamma) = \gamma(t)$ on trajectories, with $t \geq 0$. Let us also write $\mathscr{F}$ for the (smallest!) σ-algebra of $D_E[0, \infty)$ such that all π_t are measurable, and $\mathscr{F}_t$ such that all π_s for $s \leq t$ are measurable. Equipped together, the measurable space $(D_E[0, \infty), \mathscr{F})$ is the space of all paths. At last we are ready to define a Markov process.

Definition 12. *Take a family $(\mathbb{P}_x)_{x \in E}$ of probability measures on $(D_E[0, \infty), \mathscr{F})$. We say this family is a Markov process is if*

1. *for all $x \in E$ we have $\mathbb{P}_x(\gamma(0) = x) = 1$;*
2. *the map $x \mapsto \mathbb{P}_x(F)$ is measurable for all $F \in \mathscr{F}$; and*
3. *for all $s \geq 0$, $\mathbb{P}_x(\gamma(t + s) \in A \mid \mathscr{F}_t) = \mathbb{P}_{\gamma(t)}(\gamma(s) \in A)$, given any $t \geq 0$ and $A \in B(E)$.*

The last property is the most important, and is called the Markov property. Note that this "stochastic measure" is interpreted in the sense of $\mathbb{P}_x$-almost sure equality on $\mathscr{F}_t$.

In the time-based approach we feed in a time fixed t and have probabilities assigned to all points in space x. In this space-based approach, our process looks at a fixed point in space x and for each trajectory γ, a deterministic mapping $t \mapsto x$, return the probability such a trajectory will be realized under $\mathbb{P}_x$. In this sense, our index x behaves like an initial condition.

3.3 Markov Semigroups

Given some Markov process $(\mathbb{P}_x)_{x \in E}$ we define

$$\mathbb{E}_x(\Phi) = \int_{D_E[0,\infty)} \Phi \, \mathrm{d}\mathbb{P}_x$$

for $\Phi \in L^1(D_E[0,\infty), \mathscr{F}, \mathbb{P}_x)$.

As a special case, we may fix some $t \geq 0$ and take $f \in \mathscr{B}(E)$. Then, the map

$$\Phi\colon D_E[0,\infty) \to \mathbb{R} \text{ by } \gamma \mapsto f(\gamma(t))$$

is measurable with respect to $\mathscr{F}$ and absolutely-integrable with respect to any $\mathbb{P}_x$. Indeed, for any $V \in B(\mathbb{R})$ we have

$$\Phi^{-1}(V) = \{\gamma \in D_E[0,\infty) : f(\gamma(t)) \in V\} = \{\gamma \in D_E[0,\infty) : \pi_t(\gamma) \in f^{-1}(V)\}.$$

Since f is measurable we know $f^{-1}(V) \in B(E)$, and $\mathscr{F}$ is defined such that π_t is measurable. Integrability follows from the boundedness of f. In fact, such a Φ is not just L^1, but is even L^∞, which is important in the context of the following characterization of the Markov property.

Proposition 1. *Let $s, t \geq 0$. Then, for any $f \in \mathscr{B}(E)$ we have*

$$\mathbb{E}_x(f(\gamma(t+s)) \,|\, \mathscr{F}_t) = \mathbb{E}_{\gamma(t)}(f(\gamma(s))).$$

Proof. With proceed with the standard machinery. For some $A \in B(E)$ we consider the characteristic function $f \equiv \chi_A$. For for arbitrary $F \in \mathscr{F}_t$, by definition of conditional expectation we have

$$\int_F \mathbb{E}_x(f(\gamma(t+s)) \,|\, \mathscr{F}_t)\, \mathrm{d}\mathbb{P}_x(\gamma) = \int_F \chi_A(\gamma(t+s))\, \mathrm{d}\mathbb{P}_x(\gamma).$$

Since F was arbitrary, we have concluded

$$\mathbb{E}_x(\Phi(\gamma(t+s)) \,|\, \mathscr{F}_t) = \mathbb{P}_x(\gamma(t+s) \in A \,|\, \mathscr{F}_t).$$

Using the Markov property, we get that

$$\mathbb{P}_x(\gamma(t+s) \in A \,|\, \mathscr{F}_t) = \mathbb{P}_{\gamma(t)}(\gamma(s) \in A) = \mathbb{E}_{\gamma(t)}(\Phi(\gamma(s))).$$

For a simple function $f \equiv \sum_{i=1}^n a_i \chi_{A_i}$, with $a_i \in \mathbb{R}$ and $A_i \in B(E)$ pairwise disjoint, we see

$$\int_F \mathbb{E}_x(f(\gamma(t+s)) \,|\, \mathscr{F}_t)\, \mathbb{P}_x(\gamma) = \sum_{i=1}^n a_i \int_F \chi_{A_i}(\gamma(t+s))\, \mathrm{d}\mathbb{P}_x(\gamma).$$

Once more, since F was arbitrary, we conclude

$$\mathbb{E}_x(f(\gamma(t+s)) \,|\, \mathscr{F}_t) = \sum_{i=1}^n a_i \mathbb{P}_x(\gamma(t+s) \in A_i \,|\, \mathscr{F}_t) = \sum_{i=1}^n a_i \mathbb{P}_{\gamma(t)}(\gamma(s) \in A_i) = \sum_{i=1}^\infty a_i \mathbb{E}_{\gamma(t)}(\chi_{A_i}(\gamma(s))).$$

Via linearity of integration, we conclude the above is indeed $\mathbb{E}_{\gamma(t)}(f(\gamma(s)))$. For non-negative f, we approximate via simple functions, and so on (we will see an example of this procedure momentarily). $\square$

In fact, the above may even be taken as an alternative definition of the Markov property. It is generally easier to use, since we will now define a function central in analysing Markov processes in terms of this integral $\mathbb{E}_x$.

Definition 13 (Markov Semigroup)**.** *For all $t \geq 0$ define $S_t \colon \mathscr{B}(E) \to \mathscr{B}(E)$ by*

$$S_t f(x) = \mathbb{E}_x(f(\gamma(t))).$$

We say the family $(S_t)_{t\geq 0}$ is the Markov semigroup associated to $(\mathbb{P}_x)_{x \in E}$.

Some parts of this definition may be contentious, such as whether $S_t f$ really does belong to $\mathscr{B}(E)$. This is all resolved in the following proposition.

Proposition 2. *Let $(S_t)_{t\geq 0}$ be the Markov semigroup associated to a Markov process $(P_x)_{x\in E}$. Let $t, s \geq 0$ and $f, g \in \mathscr{B}(E)$. Then, S_t is linear and bounded, in particular with $\|S_t f\|_\infty \leq \|f\|_\infty$. We also have*

1. *$S_0 \equiv \mathbb{1}$;*
2. *$S_t 1 \equiv 1$; and*
3. *If $f \leq g$, then $S_t f \leq S_t g$.*

The above (in)equalities are true almost everywhere. Lastly, we have the semigroup property: $S_{t+s} \equiv S_t \circ S_s$.

Proof. Let us start by showing $S_t f \in \mathscr{B}(E)$ for $f \in \mathscr{B}(E)$. If $f \equiv \chi_A$ for some $A \in B(E)$, then

$$S_t f(x) = \int_{D_E[0,\infty)} \chi_A(\gamma(t))\, \mathrm{d}\mathbb{P}_x(\gamma) = \mathbb{P}_x(\gamma(t) \in A) = \mathbb{P}_x(\pi_t(\gamma) \in A).$$

By definition of our σ-algebra $\mathscr{F}$, we know $\pi_t^{-1}(A) \in \mathscr{F}$. So, for any Borel $V \in B(\mathbb{R})$ we have

$$S_t f^{-1}(V) = \{x \in E : \mathbb{P}_x(\pi_t(\gamma) \in A) \in V\} = \{x \in E : \mathbb{P}_x(\pi_t^{-1}(A)) \in V\},$$

which is clearly measurable since each $\mathbb{P}_x$ is a measure on the same σ-algebra $\mathscr{F}$.

Now, suppose $f \in \mathscr{B}(E)$ is non-negative. So, it may be approximated by some (pointwise!) convergent monotonic positive sequence (s_n) of simple functions. We have

$$S_t s_n(x) = \sum_{i=1}^{n} a_i^{(n)} \mathbb{P}_x\left(\pi_t(\gamma) \in A_i^{(n)}\right),$$

where $s_n = \sum_{i=1}^n a_i^{(n)} \chi_{A_i^{(n)}}$. Indeed, these too are measurable, showing that $S_t s$ is measurable for any simple s. Moving onward, since $s_n \to f$ we have

$$S_t \circ (s_n - f)(x) = \int_{D_E[0,\infty)} (s_n - f)(\gamma(t))\, \mathrm{d}\mathbb{P}_x(\gamma) \to 0.$$

This tells us $S_t s_n \to S_t f$, and so $S_t f$ itself must be measurable. This immediately extends to the mixed-sign case.

Linearity is clear by the fact that S_t is an integral. To get our bound, witness that

$$|S_t f(x)| = \left| \int_{D_E[0,\infty)} f(\gamma(t))\, \mathrm{d}\mathbb{P}_x(\gamma) \right| \leq \int_{D_E[0,\infty)} |f(\gamma(t))|\, \mathrm{d}\mathbb{P}_x(\gamma) \leq \|f\|_\infty.$$

This holds for all $x \in E$ (and really, we should sub-index our infinity norm by x, as this defines our L^∞ space).

The fact $S_t 1 \equiv 1$ follows from the definition of a probability measure. For S_0, we recall that $\mathbb{P}_x(\gamma(0) = x) = 1$ for all trajectories γ and points x.

We conclude with the semigroup property. For an arbitrary $f \in \mathscr{B}(E)$ and $t, s \geq 0$ we have

$$S_{t+s} f(x) = \mathbb{E}_x(f(\gamma(t+s))) = \mathbb{E}_x[\mathbb{E}_x(f(\gamma(t+s)) \,|\, \mathscr{F}_t)],$$

by the law of total expectation. Applying PROPOSITION 1, we have

$$\begin{aligned}\mathbb{E}_x[\mathbb{E}_x(f(\gamma(t+s))\,|\,\mathscr{F}_t)] &= \mathbb{E}_x\left[\mathbb{E}_{\gamma(t)}(f(\gamma(s)))\right] \\ &= \mathbb{E}_x[S_s f(\gamma(t))] \\ &= S_t(S_s f)(x) \\ &= (S_t \circ S_s)(f)(x),\end{aligned}$$

through two applications of the definition of the semigroup. □

So, the semigroup allows you to take a function f on our state space E, and evolve it forward by time t given some initial condition x. This is done in the sense of averaging f over the distribution of possible trajectories γ over E.

3.4 Feller Processes

Though bounded measurable functions $\mathscr{B}(E)$ are not the worst, they certainly are not the best. They are far too general a class to contain, and so we instead wish to return to $\mathscr{C}_0(E)$, those vanishing at infinity.

Definition 14 (Feller)**.** *Given a Markov semigroup $(S_t)_{t\geq 0}$, we say it is Feller if $S_t f \in \mathscr{C}_0(E)$ whenever $f \in \mathscr{C}_0(E)$. Moreover, if for any sequence $\mathscr{C}_0(E) \ni f \to 1$ we have $S_t f_n \to 1$, for all t, then we say the semigroup is conservative.*

Theorem 3. *If $(S_t)_{t\geq 0}$ is Feller, then it is strongly continuous on $\mathscr{C}_0(E)$, in the sense that for any $f \in \mathscr{C}_0(E)$ the map $t \mapsto S_t f$ is a continuous map $[0,\infty) \to \mathscr{C}_0(E)$.*

Vanishing at infinity really is crucial for the above theorem. This is as we make use of approximations that are dense only in $\mathscr{C}_0(E)$, and not for continuous functions in general.

In the previous section, we showed how we may start with any Markov process and define a semigroup, explicitly. Now, although we have introduced this Feller property, the burden remains to show it is actually realizable. This is not only true, but is so in a miraculous way.

It turns out that given some Feller semigroup (abstractly, this is just some set of time-indexed linear operators $\mathscr{C}_0(E) \to \mathscr{C}_0(E)$ with some conditions, like strong continuity and the semigroup property), we can recover the associated Markov process. This is, in fact, the Chapman-Kolmogorov master equation.

Theorem 4. *Let $(S_t)_{t\geq 0}$ be a conservative Feller semigroup, and let $s, t \geq 0$ and $x \in E$ be arbitrary. Then, there exists a probability measure $\mathbb{P}_t(x,\cdot)$ on $B(E)$ such that*

$$S_t f(x) \int_E f(y)\,\mathbb{P}_t(x,\mathrm{d}y)$$

for all $f \in \mathscr{C}_0(E)$.

Moreover, fix some $A \in B(E)$. Then, the map $x \mapsto \mathbb{P}_t(x,F)$ belongs to $\mathscr{B}(E)$. We also have the Chapman-Kolmogorov equation:

$$\mathbb{P}_{s+t}(x,A) = \int_E \mathbb{P}_s(y,A)\,\mathbb{P}_t(x,\mathrm{d}y).$$

Proof. Fix $t \geq 0$ and $x \in E$. Define the functional $f \mapsto S_t f(x)$ on $\mathscr{C}_0(E)$. By THEOREM 1 we can find a (real!) measure $\mathbb{P}_t(x,\cdot)$ such that

$$S_t f(x) = \langle \mathbb{P}_t(x,\cdot), f\rangle = \int_E f(y)\,\mathbb{P}_t(x,\mathrm{d}y).$$

By taking $f \equiv \chi_A$ for any $A \in B(E)$, we have that $f \geq 0$ and so $S_t f \geq 0$, and so we must have that $\mathbb{P}_t(x, \cdot) \geq 0$. Take now any sequence $f_n \to 1$, and observe that

$$\mathbb{P}_t(x, E) = \int_E \mathbb{P}_t(x, \mathrm{d}y) = \lim_{n\to\infty} \int_E f_n \, \mathbb{P}_t(x, \mathrm{d}y) = \lim_{n\to\infty} S_t f_n(x) = 1$$

by conservativity and dominated convergence. In all, we see that our measure $\mathbb{P}_t(x, \cdot)$ is in fact a probability measure.

For the Chapman-Kolmogorov equation, take $s, t \geq 0$. For all $f \in \mathscr{C}_0(E)$ witness that

$$\int_E f(y) \, \mathbb{P}_{s+t}(x, \mathrm{d}y) = S_{s+t} f(x) = S_s(S_t f)(x) = \int_E S_t f(y) \, \mathbb{P}_s(x, \mathrm{d}y)$$

by the semigroup property and strong continuity. Thus,

$$S_{s+t} f(x) = \int_E \left(\int_E f(z) \, \mathbb{P}_t(y, \mathrm{d}z) \right) \mathbb{P}_s(x, \mathrm{d}y).$$

Taking the specific case $f \equiv \chi_A$ for some $A \in B(E)$ we have

$$\mathbb{P}_{s+t}(x, F) = \int_E \left(\int_E \chi_A(z) \, \mathbb{P}_t(y, \mathrm{d}z) \right) \mathbb{P}_s(x, \mathrm{d}y) = \int_E \mathbb{P}_t(y, A) \, \mathbb{P}_s(x, \mathrm{d}y),$$

concluding the proof. □

We should interpret $\mathbb{P}_t(x, A)$ as the probability of a jump from $x \in E$ to somewhere in $A \subseteq E$ at time t, given our process. Then, $\mathbb{P}_t(x, \mathrm{d}y)$ is a jump to an infinitesimal distance about y. For this reason, the measures $\mathbb{P}_t(x, \cdot)$ are commonly referred to as transition probabilities (or sometimes, Markov kernels).

The Chapman-Kolmogorov equation then has the operational interpretation that a jump from x to within A at the time $s + t$ is the same as looking at all possible infinitesimal jumps x to y at time s, and then a jump from y to within A at time t.

3.5 Infinitesimal Generators

Definition 15 (Contraction Semigroup)**.** *Take a family $(S_t)_{t\geq 0}$ of continuous linear maps on a Banach space X, with $S_0 \equiv \mathbb{1}$. Suppose they form a strongly continuous semigroup, meaning $S_{s+t} = S_s S_t$ and $t \mapsto S_t x$ is a continuous map $[0, \infty) \to X$ for any $x \in X$. If we also have $\|S_t\| \leq 1$ for all $t \geq 0$, then we call the family a contraction semigroup.*

Note that Feller semigroups were an instance of a contraction semigroup. Say $f \in \mathscr{C}_0(E)$ is such that $\|f\|_\infty \leq 1$ (the optimization region relevant for the operator norm). This means $-1 \leq f(x) \leq 1$ for all $x \in E$, or that $-1 \leq f \leq 1$ (as constant functions). Since these inequalities are preserved (in particular, under any Markov semigroup), it must be that $-1 \leq S_t f \leq 1$ and so $\|S_t f\|_\infty \leq 1$.

Definition 16. *Let $(S_t)_{t\geq 0}$ be a strongly continuous semigroup. Define the operator A_S by*

$$A_S x = \lim_{t\to 0^+} \frac{S_t x - x}{t}.$$

We call A_S the infinitesimal generator of the semigroup.

Note that A_S in general is unbounded. In fact, it need not even be densely-defined! However, being a contraction semigroup is enough to guarantee not only density, but closure, which tames the discontinuity considerably.

Lemma 2. *Suppose $x \in \operatorname{dom} A_S$. Then, the derivative $\mathrm{d}(S_t x)/\mathrm{d}t$ is well-defined. Moreover,*

$$\frac{\mathrm{d}}{\mathrm{d}t} S_t x = S_t A_S x = A_S S_t x.$$

Proof. Let $h > 0$. By the semigroup property, we have

$$\frac{S_{t+h}x - S_t x}{h} = S_t\left(\frac{S_h x - x}{h}\right).$$

Taking the limit as $h \longrightarrow 0+$, we see the right-most term is merely $S_t A_S x$. Approaching from the right, we similarly start by finding,

$$\frac{S_{t-h}x - S_t x}{h} = S_{t-h}\left(\frac{x - S_h x}{h}\right).$$

Witness now that

$$\left\|S_{t-h}\left(\frac{x - S_h x}{h}\right) - S_t A_S x\right\| \leq \left\|S_{t-h}\left(\frac{x - S_h x}{h} - A_S x\right)\right\| + \|S_{t-h}A_S x - S_t A_S x\|.$$

By strong continuity, the right-most term vanishes as $h \longrightarrow 0$. Recalling then that $\|S_{t-h}\|_{\mathrm{op}} \leq 1$, we use the standard operator norm inequality to get

$$\left\|S_{t-h}\left(\frac{x - S_h x}{h} - A_S x\right)\right\| \leq \left\|\frac{x - S_h x}{h} - A_S x\right\|,$$

which also clearly vanishes in the limit. In all, we have concluded that

$$\frac{\mathrm{d}}{\mathrm{d}t} S_t x = S_t A_S x.$$

To get the last equality, it stands merely to notice that

$$\frac{S_h(S_t x) - S_t x}{h} = S_t\left(\frac{S_h x - x}{h}\right).$$

Taking $h \longrightarrow 0$, we get that $A_S S_t x = S_t A_S x$. □

Before we return to the density and closure of A_S, let us take a brief moment to recognize what the above lemma is implying, morally. It seems to be saying that, at least formally, we may interpret the infinitesimal generator as being an (the?) operator which helps (and, perhaps, even characterizes!) the semigroup (S_t) to satisfy the above differential equation. If A_S were bounded, then we could trivially solve this differential equation and find that $S_t \equiv \exp(tA_S)$.

Proposition 3. *The infinitesimal generator A_S is densely-defined and closed.*

Proof. Let $x \in X$, and for $\varepsilon > 0$ define

$$x_\varepsilon = \frac{1}{\varepsilon}\int_0^\varepsilon S_t x \,\mathrm{d}t.$$

Note this integral is a Bochner integral, and is well-defined due to strong continuity. As $\varepsilon \to 0$, due to THEOREM 2, we see that $x_\varepsilon \to x$. To show density, it then suffices to show that $x_\varepsilon \in \operatorname{dom} A_S$ for all $\varepsilon > 0$. To this end, let $t > 0$ and witness

$$\frac{1}{t}(S_t x_\varepsilon - x_\varepsilon) = \frac{1}{t\varepsilon}\left(S_t \int_0^\varepsilon S_h x \,\mathrm{d}h - \int_0^\varepsilon S_h x \,\mathrm{d}h\right).$$

Linear operators respect integration, and so

$$S_t \int_0^\varepsilon S_h x \,\mathrm{d}h = \int_t^{\varepsilon+t} S_h x \,\mathrm{d}h$$

due to the semigroup property. In all, taking the limit $t \to 0$ we see

$$\frac{1}{t}(S_t x_\varepsilon - x_\varepsilon) \to \frac{1}{\varepsilon}(S_\varepsilon x - x) = A_S x_\varepsilon,$$

once more due to Lebesgue differentiation.

Moving onto to closure, take some sequence $x_n \in \operatorname{dom} A_S$ such that $x_n \to x$ and $A_S x_n \to y$. For closure, we require that $x \in \operatorname{dom} A_S$ and that

$$y = A_S x = \lim_{t \to 0^+} \frac{S_t x - x}{t}.$$

Now, let $t > 0$. By definition of the derivative and the fact that $S_0 \equiv \mathbb{1}$, we have

$$S_t x_n - x_n = \int_0^t \frac{\mathrm{d}}{\mathrm{d}h} S_h x_n \,\mathrm{d}h = \int_0^t S_h A_S x_n \,\mathrm{d}h,$$

with the last equality following from LEMMA 2. We see that

$$\left\| \int_0^t S_h A_S x_n \,\mathrm{d}h - \int_0^t S_h y \,\mathrm{d}h \right\| \le \int_0^t \|S_h(A_S x_n - y)\| \,\mathrm{d}h \le t\|A_S x_n - y\|,$$

due to standard inequalities of the Bochner integral and the operator norm once more. Therefore, taking $n \to \infty$ tells us

$$S_t x - x = \int_0^t S_h y \,\mathrm{d}h.$$

Appealing again to Lebesgue, we divide through by t and in the limit get that

$$\frac{S_t x - x}{t} = \frac{1}{t}\int_0^t S_h y \,\mathrm{d}h \to y.$$

Therefore, we have closure. □

3.6 Markov Generators

We have now defined the infinitesimal generator, and showed that although it is unbounded, it is about as nice of an unbounded operator as we could ask for, less conditions on its adjoint. The lingering question remains, however: what to make of that differential equation?

This will be the last step in our abstraction. It does indeed turn out that each operator with the properties of an infinitesimal generator uniquely determines some contraction semigroup! The definition of an exponential via a series is insufficient for even the nicest unbounded operators (one need only look at the derivative operator, for instance). However, the limit definition

$$\exp(tA_S) = \lim_{n\to\infty}\left(\mathbb{1} - \frac{tA_S}{n}\right)^{-n}$$

is much more well-behaved, provided the resolvent set of A_S is sufficiently constrained. Indeed, this is the case.

Definition 17 (Right Half-Plane). *Denote the right half-plane*

$$\mathcal{R} = \{\lambda \in \mathbb{C} : \Re\lambda > 0\}.$$

Theorem 5. *We have the inclusion $\mathcal{R} \subseteq \rho(A_S)$. Moreover, for all $\lambda \in \mathcal{H}$ we have*

$$R_\lambda x = \int_0^\infty \exp(-\lambda t) S_t x \,\mathrm{d}t$$

and $\|R_\lambda\| \leq 1/\Re\lambda$. Here, $R_\lambda = R_\lambda(A_S)$.

Proof. Fix some $\lambda \in \mathcal{R}$. We let $t \geq 0$ and begin with the estimate

$$\|\exp(-t\lambda)S_t x\| \leq \exp(-t\Re\lambda)\|x\|.$$

In particular, this tells us $\exp(-t\lambda)S_t x$ is absolutely-integrable over $[0,\infty)$ as a function of t. From here it also follows that

$$\|R_\lambda x\| \leq \int_0^\infty \|x\| \exp(-t\Re\lambda)\,\mathrm{d}t = \frac{\|x\|}{\Re\lambda},$$

giving the desired estimate for the operator norm.

It remains to show this integral operator indeed yields the resolvent. That is, defining R_λ as this integral operator, we must show that $R_\lambda \equiv (\lambda\mathbb{1} - A_S)^{-1}$. To this end, we begin by showing $R_\lambda x \in \operatorname{dom} A_S$. Let $x \in X$ be arbitrary, and witness that for $t > 0$ we have

$$\frac{1}{t}(S_t R_\lambda x - R_\lambda x) = \frac{1}{t}\left(\int_t^\infty \exp(-\lambda(h-t))S_h x\,\mathrm{d}h - \int_0^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h\right)$$

through the semigroup property. We write this first integral as

$$\frac{1}{t}\int_t^\infty \exp(-\lambda(h-t))S_h x\,\mathrm{d}h = \frac{\exp(\lambda t)-1}{t}\int_t^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h + \frac{1}{t}\int_t^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h,$$

where we add and subtract 1. In all, we are left to evaluate

$$\frac{\exp(\lambda t)-1}{t}\int_t^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h + \frac{1}{t}\left(\int_t^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h - \int_0^\infty \exp(-\lambda h)S_h x\,\mathrm{d}h\right).$$

Taking the limit, we see the first term approaches

$$\lambda \int_0^\infty \exp(-\lambda h) S_h x \, \mathrm{d}h = \lambda R_\lambda x,$$

while for the last term we have

$$\frac{1}{t}\left(\int_t^\infty \exp(-\lambda h) S_h x \, \mathrm{d}h - \int_0^\infty \exp(-\lambda h) S_h x \, \mathrm{d}h\right) = \frac{-1}{t} \int_0^t \exp(-\lambda h) S_h x \, \mathrm{d}h \to -x.$$

In all, we have $A_S R_\lambda x = \lambda R_\lambda x - x$. This simultaneously shows that $R_\lambda x \in \operatorname{dom} A_S$ and that $(\lambda \mathbb{1} - A_S) R_\lambda x = x$. To get the other inverse, we notice that

$$R_\lambda (\lambda \mathbb{1} - A_S) x = \lambda \int_0^\infty \exp(-\lambda h) S_h x \, \mathrm{d}h - \int_0^\infty \exp(-\lambda h) S_h A_S x \, \mathrm{d}h.$$

From LEMMA 2 we recognize that

$$\begin{aligned}\int_0^\infty \exp(-\lambda h) S_h A_S x \, \mathrm{d}h &= \int_0^\infty \exp(-\lambda h) \frac{\mathrm{d}}{\mathrm{d}h} S_h x \, \mathrm{d}h \\ &= \exp(-\lambda h) S_h x \Big|_{h=0}^{\infty} + \lambda \int_0^\infty \exp(-\lambda h) S_h x \, \mathrm{d}h \\ &= -x + \lambda R_\lambda x\end{aligned}$$

via integration by parts, whence the conclusion follows. □

This lets us take our final leap into the abstract, finding that these properties of our constructed infinitesimal generator are sufficient for us to take them as a starting point, and conversely construct the semigroup. This is the content of the following theorem.

Theorem 6 (Hille-Yosida). *Let A be a densely-defined closed operator such that $\mathcal{R} \subseteq \rho(A)$ and*

$$\|R_\lambda(A)\| \le \frac{1}{\Re \lambda}$$

Define $A_n = nAR_n(A)$ for $n > 0$ and for $t \ge 0$ define

$$e^{tA} = \lim_{n\to\infty} \exp(tA_n).$$

Then, $(e^{tA})_{t\ge 0}$ is a contraction semigroup, uniquely with A as its infinitesimal generator.

Some remarks are in order. First, the notation $\exp(tA)$ is purely formal, since again we cannot define the exponential of A in general. However, $A_n \to A$ (as operators restricted to $\operatorname{dom} A$), and the exponential of A_n is perfectly well-defined (indeed, A_n is always bounded). In the event that A is bounded, then of course its exponential is no longer formal, and actually is a true equality. Lastly, we will now, in general, write $e^{tA} = S_t$ and $A = A_S$ for contraction semigroups in general, implicitly inserting the above remarks.

There is yet still a subtle loose end, that of Feller semigroups. These are instances of contraction semigroups, yes, and so Hille-Yosida does characterize – almost entirely – their infinitesimal generators. However, Feller semigroups must preserve positivity as well. In fact, all Markov semigroups must! We need an extra condition on our infinitesimal generator to ensure this happens.

Definition 18 (Dissipative)**.** *We say a (perhaps unbounded!) operator A is dissipative if*

$$\|\lambda x - Ax\| \geq \lambda \|x\|$$

for all $\lambda > 0$ and $x \in \operatorname{dom} A$.

Proposition 4. *Let $(e^{tA})_{t\geq 0}$ be a Feller semigroup on $\mathscr{C}_0(E)$. Then, A is dissipative.*

Proof. Let $f \in \mathscr{C}_0(E)$. Due to LEMMA 1 we know it attains its maximum at some $x_0 \in E$. Without loss of generality assume $f(x_0) \geq 0$. We know that

$$Af(x_0) = \lim_{t\to 0^+} \frac{e^{tA} f(x_0) - f(x_0)}{t}.$$

Now, let $f^+ = \max\{f, 0\}$. Since Feller semigroups preserve positivity and are contractions, for all $t \geq 0$ we have

$$S_t f \leq S_t f^+ \leq \|S_t f^+\| \leq \|f^+\| = f(x_0).$$

Thus,

$$Af(x_0) \leq \lim_{t\to 0^+} \frac{f(x_0) - f(x_0)}{t} = 0.$$

Now, let $\lambda > 0$, and by definition of the infinity-norm we have

$$\|\lambda f - Af\| \geq \|\lambda x_0 - Af(x_0)| = \lambda f(x_0) - Af(x_0) \geq \lambda f(x_0) = \lambda \|f\|,$$

since $Af(x_0) \leq 0 \leq \lambda f(x_0)$. □

So, Feller semigroups have dissipative infinitesimal generators. The structure of our narrative implies the reverse holds as well. Indeed, this is true.

Theorem 7 (Lumer-Phillips)**.** *Let A be a dissipative operator such that $\lambda \mathbb{1} - A$ is surjective for some $\lambda > 0$. Then, A is the infinitesimal generator of a contraction semigroup.*

This is, more or less, a restatement of Hille-Yosida. First, the dissipation of A gives the estimate necessary for the norm of the resolvent. Second, a non-empty resolvent set is enough to imply the closure of an operator (essentially just limit of the sequence on the resolvent). We this all in mind, we may now posit the following definition – that an operator satisfying the hypotheses of the Lumer-Phillips theorem is a Markov generator. In the specific case that A is an operator over $\mathscr{C}_0(\mathbb{R})$, it will be the generator of a Feller process.

3.7 Weiner Process

Let us conclude with a simple example. We know what the infinitesimal generator of a Weiner process on $E = \mathbb{R}$ ought to be:

$$A = \frac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}x^2}.$$

We consider this to be an unbounded operator on $\mathscr{C}_0(\mathbb{R})$. Of course, for such an operator to make sense we must (at least!) restrict ourselves to twice continuously-differentiable functions. That is,

$$\operatorname{dom} A = \mathscr{C}_0(\mathbb{R}) \cap C^2(\mathbb{R}).$$

Fortunately, it is a standard result that this domain is dense.

Getting dissipativity is not hard: in the proof of PROPOSITION 4 we deduced that this is implied if for all $f \in \operatorname{dom} A$ which have a positive maximum, say at x_0, then $Af(x_0) \leq 0$. However, we know this already due to simple calculus! Indeed, if f has a local maximum at x_0, then its second derivative at x_0 is always negative.

The difficult part is showing it holds a non-empty resolvent set. For this, let $\lambda > 0$ and take some $g \in \mathscr{C}_0(\mathbb{R})$. What this amounts to is finding some unique $f \in \operatorname{dom} A$ such that

$$g = \lambda f - Af = \lambda f - \frac{1}{2} f''.$$

The existence and uniqueness of such an f proves that the resolvent operator is invertible. Of course, we are free to pick λ (or rather, fix any positive one), but g must be a generic function which vanishes at infinity.

It is possible to show that this differential equation is generically solvable, with solutions taking the form

$$f(x) = c_1 \exp\left(-\sqrt{2\lambda} x\right) + c_2 \exp\left(\sqrt{2\lambda} x\right) + \frac{2}{\sqrt{2\lambda}} \int_0^x g(y) \sinh\left(\sqrt{2\lambda}(x-y)\right) \mathrm{d}y,$$

with constants c_1 and c_2. Rewriting the hyperbolic sine, we find this is equal to

$$\begin{aligned} f(x) = \left[c_1 + \frac{1}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(-\sqrt{2\lambda} y\right)\right] \exp\left(\sqrt{2\lambda} x\right) \\ + \left[c_2 - \frac{1}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(\sqrt{2\lambda} y\right) \mathrm{d}y\right] \exp\left(-\sqrt{2\lambda} x\right). \end{aligned}$$

Suppose now that $f \in \mathscr{C}_0(\mathbb{R})$. Indeed, as $x \to \infty$ we have

$$\frac{\exp\left(-\sqrt{2\lambda} x\right)}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(\sqrt{2\lambda} y\right) \mathrm{d}y = \frac{g(x)}{\sqrt{2\lambda}} \to 0$$

since $g \in \mathscr{C}_0(\mathbb{R})$ by assumption. The c_2 vanishes of course as well. Moving to the remaining term, we examine the parenthetical component:

$$\left| c_1 + \frac{1}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(-\sqrt{2\lambda} y\right) \right| \leq |c_1| + \|g\| \int_0^x \exp\left(-\sqrt{2\lambda} y\right) \mathrm{d}y.$$

Taking $x \to \infty$, this integral converges. So, as f vanishes at infinity, we must have that

$$c_1 = -\frac{1}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(-\sqrt{2\lambda} y\right)$$

in order for the remaining term to disappear. Likewise, taking $x \to -\infty$ will show that

$$c_2 = \frac{1}{\sqrt{2\lambda}} \int_0^x g(y) \exp\left(\sqrt{2\lambda} y\right) \mathrm{d}y.$$

Therefore, such an f is unique. This shows that A indeed satisfies the hypotheses necessary to be a Markov generator. As a consequence, if we prove that A does generate a Weiner process, we immediately know too that the Weiner process is Feller.

References

[Sha48] Claude Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27 (1948).

[SR80] Barry Simon and Michael Reed. *Methods of Modern Mathematical Physics. Functional Analysis*. Academic Press, 1980.

[Lat04] Bruno Latour. *Politics of Nature. How to Bring the Sciences into Democracy*. Harvard University Press, 2004.

[Shr04] Steven Shreve. *Stochastic Calculus for Finance II*. Springer, 2004.

[Bre10] Haim Brezis. *Functional Analysis, Sobolev Spaces, and Partial Differential Equations*. Springer, 2010.

[Hay13] Brian Hayes. "First Links in the Markov Chain". In: *American Scientist* 101 (2 2013).

[Sco13] Matt Scott. *Applied Stochastic Processes in Science and Engineering*. University of Waterloo, 2013.

[Hun14] John Hunter. *Notes on Partial Differential Equations*. 2014. URL: `math.ucdavis.edu/~hunter/pdes/pdes.html`.

[MRT18] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, 2018.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2019.

[Gui20] Paolo Guiotto. *Markov Semigroups*. 2020. URL: `math.unipd.it/~daipra/didattica/Bologna12/MarkovSemgroups-12.pdf`.

[Bas22] Richard Bass. *Real Analysis for Graduate Students*. 2022. URL: `bass.math.uconn.edu/real.html`.