Chapter 14

Feller Processes

Section 14.1 fulfills the demand, made last time, for an example of a Markov process which is not strongly Markovian.

Section 14.2 makes explicit the idea that the transition kernels of a Markov process induce a kernel over sample paths, mostly to fix notation for later use.

Section 14.3 defines Feller processes, which link the cadlag and strong Markov properties.

14.1 An Example of a Markov Process Which Is Not Strongly Markovian

This is taken from Fristedt and Gray (1997, pp. 626–627).

Example 138 *We will construct an $\mathbb{R}^2$-valued Markov process on $[0, \infty)$ which is not strongly Markovian. Begin by defining the following map from $\mathbb{R}$ to $\mathbb{R}^2$:*

$$f(w) = \begin{cases} (w, 0) & w \leq 0 \\ (\sin w, 1 - \cos w) & 0 < w < 2\pi \\ (w - 2\pi, 0) & w \geq 2\pi \end{cases} \tag{14.1}$$

When w is less than zero or above 2π, $f(w)$ moves along the x axis of the plane; in between, it moves along a circle of radius 1, centered at $(0, 1)$, which it enters and leaves at the origin. Notice that f is invertible everywhere except at the origin, which is ambiguous between $w = 0$ and $w = 2\pi$.

Let $X(t) = f(W(t) + \pi)$, where $W(t)$ is a standard Wiener process. At all t, $\mathbb{P}(W(t) + \pi = 0) = \mathbb{P}(W(t) + \pi = 2\pi) = 0$, so, with probability 1, $X(t)$ can be inverted to get $W(t)$. Since $W(t)$ is a Markov process, it follows that $\mathbb{P}(X(t+h) \in B | X(t) = x) = \mathbb{P}\left(X(t+h) \in B | \mathcal{F}_t^X\right)$ almost surely, i.e., X is Markov. Now consider $\tau = \inf_t X(t) = (0,0)$, the hitting time of the origin. This is clearly an F^X-optional time, and equally clearly almost surely finite, because, with probability 1, $W(t)$ will leave the interval $(-\pi, \pi)$ within a finite *time. But, equally clearly, the future behavior of X will be very different if it hits the origin because $W = \pi$ or because $W = -\pi$, which cannot be determined* just *from X. Hence, there is at least one optional time at which X is not strongly Markovian, so X is not a strong Markov process.*

14.2 Markov Families

We have been fairly cavalier about the idea of a Markov process having a particular initial state or initial distribution, basically relying on our familiarity with these ideas from elementary courses on stochastic processes. For future purposes, however, it is helpful to bring this notions formally within our general framework, and to fix some notation.

Definition 139 (Initial Distribution, Initial State) *Let Ξ be a Borel space with σ-field $\mathcal{X}$, T be a one-sided index set, and $\mu_{t,s}$ be a collection of Markovian transition kernels on Ξ. Then the* Markov process with initial distribution ν, *X_ν, is the Markov process whose finite-dimensional distributions are given by the action of $\mu_{t,s}$ on ν. That is, for all $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$,*

$$X_\nu(0), X_\nu(t_1), X_\nu(t_2), \ldots X_\nu(t_n) \quad \sim \quad \nu \otimes \mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \ldots \otimes \mu_{t_{n-1},t_n} \tag{14.2}$$

If $\nu = \delta(x - a)$, the delta distribution at a, then we write X_a and call it the Markov process with initial state *a.*

The existence of processes with given initial distributions and initial states is a trivial consequence of Theorem 103, our general existence result for Markov processes.

Lemma 140 *For every initial state x, there is a probability distribution P_x on $\Xi^T, \mathcal{X}^T$. The function $P_x(A) : \Xi \times \mathcal{X}^T \to [0, 1]$ is a probability kernel.*

PROOF: The initial state fixes all the finite-dimensional distributions, so the existence of the probability distribution follows from Theorem 23. The fact that $P_x(A)$ is a kernel is a straightforward application of the definition of kernels (Definition 30). □

Definition 141 *The* Markov family *corresponding to a given set of transition kernels $\mu_{t,s}$ is the collection of all P_x.*

That is, rather than thinking of a different stochastic process for each initial state, we can simply think of different distributions over the path space Ξ^T. This suggests the following definition.

Definition 142 *For a given initial distribution ν on Ξ, we define a distribution on the paths in a Markov family as, $\forall A \in \mathcal{X}^T$,*

$$P_\nu(A) \quad \equiv \quad \int_\Xi P_x(A)\nu(dx) \tag{14.3}$$

In physical contexts, we sometimes refer to distributions ν as *mixed states*, as opposed to the *pure states* x, because the path-space distributions induced by the former are mixtures of the distributions induced by the latter. You should check that the distribution over paths given by a Markov process with initial distribution ν, according to Definition 139, agrees with that given by Definition 142.

14.3 Feller Processes

Working in the early 1950s, Feller showed that, by imposing very reasonable conditions on the semi-group of evolution operators corresponding to a homogeneous Markov process, one could obtain very powerful results about the near-continuity of sample paths (namely, the existence of cadlag versions), about the strong Markov property, etc. Ever since, processes with such nice semi-groups have been known as *Feller processes*, or sometimes as *Feller-Dynkin processes*, in recognition of Dynkin's work in extending Feller's original approach. Unfortunately, to first order there are as many definitions of a Feller semi-group as there are books on Markov processes. I am going to try to follow Kallenberg as closely as possible, because his version is pretty clearly motivated, and you've already got it.

One point to notice is that, in developing the theory of Feller operators, we need to switch from operators on L_1, where we have been working before, to operators on L_∞. The L_∞ norm, $\sup_x |f(x)|$, is much stronger than the L_1 norm, $\int |f(x)|\mu(dx)$, and the former will let us make some regularity arguments which just aren't possible in the latter, at least not without a lot of extra machinery and assumptions.

As usual, we warm up with some definitions.

Definition 143 (Positive Operator) *An operator O is* positive *when $f \geq 0$ a.e. implies $Of \geq 0$ a.e.*

Definition 144 (Contraction Operator) *An operator O is an L_p-*contraction *when $\|Of\|_p \leq \|f\|_p$.*

Definition 145 (Strongly Continuous Semigroup) *A semigroup of operators O_t is* strongly continuous in the L_p sense *on a set of functions L when, $\forall f \in L$*

$$\lim_{t \to 0} \|O_t f - f\|_p = 0 \tag{14.4}$$

In the two preceding definitions, the p in L_p should be understood to be anything from 1 to ∞ inclusive.

Definition 146 (Conservative Operator) *An operator O is* conservative *when $O\mathbf{1}_\Xi = \mathbf{1}_\Xi$.*

In these terms, our earlier Markov operators are linear, positive, conservative L_1 contractions.

Lemma 147 *If O_t is a strongly continuous semigroup of linear L_p contractions, then, for each f, $O_t f$ is a continuous function of t.*

Proof: Continuity here means that $\lim_{t' \to t} \|O_{t'} f - O_t\|_p = 0$ — we are using the L_p norm as our metric in function space. Consider first the limit from above:

$$\begin{aligned} \|O_{t+h} f - O_t f\|_p &= \|O_t(O_h f - f)\|_p && (14.5) \\ &\leq \|O_h f - f\|_p && (14.6) \end{aligned}$$

since the operators are contractions. Because they are strongly continuous, $\|O_h f - f\|_p$ can be made smaller than any $\epsilon > 0$ by taking h sufficiently small. Hence $\lim_{h \downarrow 0} O_{t+h} f$ exists and is $O_t f$. Similarly, for the limit from below,

$$\begin{aligned} \|O_{t-h} f - O_t f\|_p &= \|O_t f - O_{t-h} f\|_p && (14.7) \\ &= \|O_{t-h}(O_h f - f)\|_p && (14.8) \\ &\leq \|O_h f - f\|_p && (14.9) \end{aligned}$$

using the contraction property again. So $\lim_{h \downarrow 0} O_{t-h} f = O_t f$, also, and we can just say that $\lim_{t' \to t} O_{t'} f = O_t f$. □

Remark: The result actually holds if we just assume strong continuity, without contraction, but the proof isn't so pretty; see Ethier and Kurtz (1986, ch. 1, corollary 1.2, p. 7).

Definition 148 (Feller Semigroup) *A semigroup of linear, positive, conservative L_∞ contraction operators K_t is a* Feller semigroup *if, for every $f \in C_0$ and $x \in \Xi$, (Definition 127),*

$$\begin{aligned} K_t f &\in C_0 && (14.10) \\ \lim_{t \to 0} K_t f(x) &= f(x) && (14.11) \end{aligned}$$

Remark: Some authors omit the requirement that K_t be conservative. Also, this is just the homogeneous case, and one can define inhomogeneous Feller semigroups. However, the homogeneous case will be plenty of work enough for us!

Definition 149 (Feller Process) *A homogeneous Markov family X is a* Feller process *when, for all $x \in \Xi$,*

$$\begin{aligned} \forall t,\ y \to x \quad &\Rightarrow \quad X_y(t) \xrightarrow{d} X_x(t) && (14.12) \\ t \to 0 \quad &\Rightarrow \quad X_x(t) \xrightarrow{P} x && (14.13) \end{aligned}$$

Lemma 150 *Eq. 14.10 holds if and only if Eq. 14.12 does.*

PROOF: Exercise 14.2. □

Lemma 151 *Eq. 14.11 holds if and only if Eq. 14.13 does.*

PROOF: Exercise 14.3. □

Theorem 152 *A Markov process is a Feller process if and only if its evolution operators form a Feller semigroup.*

PROOF: Combine the lemmas. □

Feller semigroups in continuous time have generators, as in Chapter 12. In fact, the generator is *especially* useful for Feller semigroups, as seen by this theorem.

Theorem 153 (Generator of a Feller Semigroup) *If K_t and H_t are Feller semigroups with generator G, then $K_t = H_t$.*

PROOF: Because Feller semigroups consist of contractions, the Hille-Yosida Theorem 130 applies, and, for every positive λ, the resolvent $R_\lambda = (\lambda I - G)^{-1}$. Hence, if K_t and H_t have the same generator, they have the same resolvent operators. But this means that, for every $f \in C_0$ and x, $K_t f(x)$ and $H_t f(x)$ have the same Laplace transforms. Since, by Eq. 14.11 $K_t f(x)$ and $H_t f(x)$ are both right-continuous, their Laplace transforms are unique, so $K_t f(x) = H_t f(x)$. □

Theorem 154 *Every Feller semigroup K_t with generator G is strongly continuous on* $\mathrm{Dom}(G)$.

PROOF: From Corollary 126, we have, as seen in Chapter 13, for all $t \geq 0$,

$$K_t f - f = \int_0^t K_s G f ds \quad (14.14)$$

Clearly, the right-hand side goes to zero as $t \to 0$. □

The two most important properties of Feller processes is that they are cadlag (or, rather, always have cadlag versions), and that they are strongly Markovian. First, let's look at the cadlag property. We need a result which I really should have put in Chapter 8.

Proposition 155 *Let Ξ be a locally compact, separable metric space with metric ρ, and let X be a separable Ξ-valued stochastic process on T. For given $\epsilon, \delta > 0$, define $\alpha(\epsilon, \delta)$ to be*

$$\inf_{\Gamma \in \mathcal{F}_s^X:\ \mathbb{P}(\Gamma)=1} \sup_{s,t \in T:\ s \leq t \leq s+\delta} \mathbb{P}\left(\omega :\ \rho(X(s,\omega), X(t,\omega)) \geq \epsilon,\ \omega \in \Gamma | \mathcal{F}_s^X\right) \quad (14.15)$$

If, for all ϵ,

$$\lim_{\delta \to 0} \alpha(\epsilon, \delta) = 0 \quad (14.16)$$

then X has a cadlag version.

Proof: Combine Theorem 2 and Theorem 3 of Gikhman and Skorokhod (1965/1969, Chapter IV, Section 4). □

Lemma 156 *Let X be a separable homogeneous Markov process. Define*

$$\alpha(\epsilon, \delta) = \sup_{t \in T:\ 0 \leq t \leq \delta;\ x \in \Xi} \mathbb{P}\left(\rho(X_x(t), x) \geq \epsilon\right) \tag{14.17}$$

If, for every $\epsilon > 0$,

$$\lim_{\delta \rightarrow 0} \alpha(\epsilon, \delta) = 0 \tag{14.18}$$

then X has a cadlag version.

Proof: The α in this lemma is clearly the α in the preceding proposition, using the fact that X is Markovian with respect to its natural filtration and homogeneous. □

Lemma 157 *A separable homogeneous Markov process X has a cadlag version if*

$$\lim_{\delta \downarrow 0} \sup_{x \in \Xi,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(t), x)\right] = 0 \tag{14.19}$$

Proof: Start with the Markov inequality.

$$\forall x, t > 0, \epsilon > 0,\ \mathbb{P}\left(\rho(X_x(t), x) \geq \epsilon\right) \leq \frac{\mathbf{E}\left[\rho(X_x(t), x)\right]}{\epsilon} \tag{14.20}$$

$$\forall x, \delta > 0, \epsilon > 0,\ \sup_{0 \leq t \leq \delta} \mathbb{P}\left(\rho(X_x(t), x) \geq \epsilon\right) \leq \sup_{0 \leq t \leq \delta} \frac{\mathbf{E}\left[\rho(X_x(t), x)\right]}{\epsilon} \tag{14.21}$$

$$\forall \delta > 0, \epsilon > 0,\ \sup_{x,\ 0 \leq t \leq \delta} \mathbb{P}\left(\rho(X_x(\delta), x) \geq \epsilon\right) \leq \frac{1}{\epsilon} \sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(\delta), x)\right] \tag{14.22}$$

Taking the limit as $\delta \downarrow 0$, we have, for all $\epsilon > 0$,

$$\lim_{\delta \rightarrow 0} \alpha(\epsilon, \delta) \leq \frac{1}{\epsilon} \lim_{\delta \downarrow 0} \sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(\delta), x)\right] = 0 \tag{14.23}$$

So the preceding lemma applies. □

Theorem 158 (Feller Implies Cadlag) *Every Feller process X has a cadlag version.*

Proof: First, by the usual arguments, we can get a separable version of X. Next, we want to show that the last lemma is satisfied. Notice that, because Ξ is compact, $\lim_x \rho(x_n, x) = 0$ if and only if $f_k(x_n) \rightarrow f_k(x)$, for all f_k in some countable dense subset of the continuous functions on the state space.[1] Since the Feller semigroup is strongly continuous on the domain of its generator (Theorem 154), and that domain is dense in C_0 by the Hille-Yosida Theorem (130), we can pick our f_k to be in this class. The strong continuity is with respect to the L_∞ norm, so $\sup_x |K_t f(x) - K_s f(x)| = \sup_x |K_s(K_{t-s} f(x) - f(x))| \to 0$ as $t - s \to 0$, for every $f \in C_0$. But $\sup_x |K_t f(x) - K_s f(x)| = \sup_x \mathbf{E}\left[|f(X_x(t)) - f(X_x(s))|\right]$. So $\sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[|f(X_x(t)) - f(x)|\right] \to 0$ as $\delta \to 0$. Now Lemma 157 applies. □

Remark: Kallenberg (Theorem 19.15, p. 379) gives a different proof, using the existence of cadlag paths for certain kinds of supermartingales, which he builds using the resolvent operator. This seems to be the favored approach among modern authors, but obscures, somewhat, the work which the Feller properties do in getting the conclusion.

Theorem 159 (Feller Processes are Strongly Markovian) *Any Feller process X is strongly Markovian with respect to $\mathcal{F}^{X+}$, the right-continuous version of its natural filtration.*

PROOF: The strong Markov property holds if and only if, for all bounded, continuous functions f, $t \geq 0$ and $\mathcal{F}^{X+}$-optional times τ,

$$\mathbf{E}\left[f(X(\tau + t)) | \mathcal{F}_\tau^{X+}\right] = K_t f(X(\tau)) \tag{14.24}$$

We'll show this holds for arbitrary, fixed choices of f, t and τ. First, we discretize time, to exploit the fact that the Markov and strong Markov properties coincide for discrete parameter processes. For every $h > 0$, set

$$\tau_h \equiv \inf_u \{u \geq \tau :\ u = kh,\ k \in \mathbb{N}\} \tag{14.25}$$

Now τ_h is almost surely finite (because τ is), and $\tau_h \to \tau$ a.s. as $h \to 0$. We construct the discrete-parameter sequence $X_h(n) = X(nh)$, $n \in \mathbb{N}$. This is a Markov sequence with respect to the natural filtration, i.e., for every bounded continuous f and $m \in \mathbb{N}$,

$$\mathbf{E}\left[f(X_h(n + m)) | \mathcal{F}_n^X\right] = K_{mh} f(X_h(n)) \tag{14.26}$$

Since the Markov and strong Markov properties coincide for Markov sequences, we can now assert that

$$\mathbf{E}\left[f(X(\tau_h + mh)) | \mathcal{F}_{\tau_h}^X\right] = K_{mh} f(X(\tau_h)) \tag{14.27}$$

Since $\tau_h \geq \tau$, $\mathcal{F}_\tau^X \subseteq \mathcal{F}_{\tau_h}^X$. Now pick any set $B \in \mathcal{F}_\tau^{X+}$ and use smoothing:

$$\mathbf{E}\left[f(X(\tau_h + t))\mathbf{1}_B\right] = \mathbf{E}\left[K_t f(X(\tau_h))\mathbf{1}_B\right] \tag{14.28}$$

$$\mathbf{E}\left[f(X(\tau + t))\mathbf{1}_B\right] = \mathbf{E}\left[K_t f(X(\tau))\mathbf{1}_B\right] \tag{14.29}$$

[1] Roughly speaking, if $f(x_n) \to f(x)$ for *all* continuous functions f, it should be obvious that there is no way to avoid having $x_n \to x$. Picking a countable dense subset of functions is still enough.

where we let $h \downarrow 0$, and invoke the fact that $X(t)$ is right-continuous (Theorem 158) and $K_t f$ is continuous. Since this holds for arbitrary $B \in \mathcal{F}_\tau^{X+}$, and $K_t f(X(\tau))$ has to be $\mathcal{F}_\tau^{X+}$-measurable, we have that

$$\mathbf{E}\left[f(X(\tau+t))|\mathcal{F}_\tau^{X+}\right] = K_t f(X(\tau)) \tag{14.30}$$

as required. □

Here is a useful consequence of Feller property, related to the martingale-problem properties we saw last time.

Theorem 160 (Dynkin's Formula) *Let X be a Feller process with generator G. Let α and β be two almost-surely-finite $\mathcal{F}$-optional times, $\alpha \leq \beta$. Then, for every continuous $f \in \mathrm{Dom}(G)$,*

$$\mathbf{E}\left[f(X(\beta)) - f(X(\alpha))\right] = \mathbf{E}\left[\int_\alpha^\beta Gf(X(t))dt\right] \tag{14.31}$$

Proof: Exercise 14.4. □

Remark: A large number of results very similar to Eq. 14.31 are *also* called "Dynkin's formula". For instance, Rogers and Williams (1994, ch. III, sec. 10, pp. 253–254) give that name to *three* different equations. Be careful about what people mean!

14.4 Exercises

Exercise 14.1 (Yet Another Interpretation of the Resolvents) *Consider again a homogeneous Markov process with transition kernel μ_t. Let τ be an exponentially-distributed random variable with rate λ, independent of X. Show that $\mathbf{E}[K_\tau f(x)] = \lambda R_\lambda f(x)$.*

Exercise 14.2 (The First Pair of Feller Properties) *Prove Lemma 150. Hint: you may use the fact that, for measures, $\nu_t \to \nu$ if and only if $\nu_t f \to \nu f$, for every bounded, continuous f.*

Exercise 14.3 (The Second Pair of Feller Properties) *Prove Lemma 151.*

Exercise 14.4 (Dynkin's Formula) *Prove Theorem 160*