Chapter 14

Feller Processes

Section 14.1 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 14.2 defines Feller processes, which link the cadlag and strong Markov properties.

14.1 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 173 (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 $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$,*

$$X_\nu(0), X_\nu(t_1), X_\nu(t_2), \ldots X_\nu(t_n) \quad \sim \quad \nu \otimes \mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \ldots \otimes \mu_{t_{n-1},t_n} \tag{14.1}$$

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 106, our general existence result for Markov processes.

Lemma 174 (Kernel from Initial States to Paths) *For every initial state x, there is a probability distribution P_x on $\Xi^T, \mathcal{X}^T$. The function $P_x(A) : \Xi \times \mathcal{X}^T \rightarrow [0,1]$ 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 175 (Markov Family) *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 176 (Mixtures of Path Distributions (Mixed States)) *For a given initial distribution* ν *on* Ξ*, we define a distribution on the paths in a Markov family as,* $\forall A \in \mathcal{X}^T$,

$$P_\nu(A) \equiv \int_\Xi P_x(A)\nu(dx) \tag{14.2}$$

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 173, agrees with that given by Definition 176.

14.2 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.

Definition 177 (Feller Process) *A continuous-time homogeneous Markov family* X *is a* Feller process *when, for all* $x \in \Xi$,

$$\forall t,\ y \to x \quad \Rightarrow \quad X_y(t) \xrightarrow{d} X_x(t) \tag{14.3}$$

$$t \to 0 \quad \Rightarrow \quad X_x(t) \xrightarrow{P} x \tag{14.4}$$

Remark 1: The first property basically says that the dynamics are a smooth function of the initial state. Recall[1] that if we have an ordinary differential equation, $dx/dt = f(x)$, and the function f is reasonably well-behaved, the existence and uniqueness theorem tells us that there is a function $x(t, x_0)$ satisfying the equation, and such that $x(0, x_0) = x_0$. Moreover, $x(t, x_0)$ is a continuous function of x_0 for all t. The first Feller property is one counterpart of this for stochastic processes. This is a very natural assumption to make in physical or social modeling, that very similar initial conditions should lead to very similar developments.

Remark 2: The second property imposes a certain amount of smoothness on the trajectories themselves, and not just how they depend on the initial state. It's a pretty well-attested fact about the world that teleportation does not take place, and that its state changes in a reasonably continuous manner ("*natura non facit saltum*", as they used to say). However, the second Feller property does *not* require that the sample paths actually be continuous. We will see below that they are, in general, merely cadlag. So a certain limited amount of teleportation, or *salti*, is allowed after all. We do not think this actually happens, but it is a convenience when using a discrete set of states to approximate a continuous variable.

In developing the theory of Feller processes, we will work mostly with the time-evolution operators, acting on observables, rather than the Markov operators, acting on distributions. This is traditional in this part of the theory, as it seems to let us get away with less technical machinery, in effect because the norm $\sup_x |f(x)|$ is stronger than the norm $\int |f(x)|dx$. Of course, because the two kinds of operators are adjoint, you *can* work out everything for the Markov operators, if you want to.

As usual, we warm up with some definitions. The first two apply to operators on any normed linear space L, which norm we generically write as $\|\cdot\|$. The second two apply specifically when L is a space of real-valued functions on some Ξ, such as L_p, p from 1 to ∞ inclusive.

Definition 178 (Contraction Operator) *An operator A is an L-*contraction *when $\|Af\| \leq \|f\|$.*

Definition 179 (Strongly Continuous Semigroup) *A semigroup of operators A_t is* strongly continuous in the L sense *on a set of functions D when, $\forall f \in D$*

$$\lim_{t \to 0} \|A_t f - f\| = 0 \tag{14.5}$$

Definition 180 (Positive Operator) *An operator A on a function space L is* positive *when $f \geq 0$ a.e. implies $Af \geq 0$ a.e.*

[1] Or read Arnol'd (1973), if memory fails you.

Definition 181 (Conservative Operator) *An operator A is* conservative *when $A\mathbf{1}_\Xi = \mathbf{1}_\Xi$.*

In these terms, our earlier Markov operators are linear, positive, conservative contractions, either on $L_1(\mu)$ (for densities) or $\mathcal{M}(\Xi)$ (for measures).

Lemma 182 (Continuous semi-groups produce continuous paths in function space) *If A_t is a strongly continuous semigroup of linear contractions on L, then, for each $f \in L$, $A_t f$ is a continuous function of t.*

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

$$\begin{aligned} \|A_{t+h} f - A_t f\| &= \|A_t(A_h f - f)\| &(14.6)\\ &\leq \|\|O_h f - f\| &(14.7) \end{aligned}$$

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

$$\begin{aligned} \|A_{t-h} f - A_t f\| &= \|A_t f - A_{t-h} f\| &(14.8)\\ &= \|A_{t-h}(A_h f - f)\| &(14.9)\\ &\leq \|A_h f - f\| &(14.10) \end{aligned}$$

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

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).

There is one particular function space L we will find especially interesting.

Definition 183 (Continuous Functions Vanishing at Infinity) *Let Ξ be a locally compact and separable metric space. The class of functions C_0 will consist of functions $f : \Xi \mapsto R$ which are continuous and for which $\|x\| \to \infty$ implies $f(x) \to 0$. The norm on C_0 is $\sup_x |f(x)|$.*

Definition 184 (Feller Semigroup) *A semigroup of linear, positive, conservative contraction operators K_t is a* Feller semigroup *if, for every $f \in C_0$ and $x \in \Xi$, (Definition 183),*

$$\begin{aligned} K_t f &\in C_0 &(14.11)\\ \lim_{t \to 0} K_t f(x) &= f(x) &(14.12) \end{aligned}$$

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!

You can guess how Feller semi-groups relate to Feller processes.

Lemma 185 (The First Pair of Feller Properties) *Eq. 14.11 holds if and only if Eq. 14.3 does.*

PROOF: Exercise 14.2. □

Lemma 186 (The Second Pair of Feller Properties) *Eq. 14.12 holds if and only if Eq. 14.4 does.*

PROOF: Exercise 14.3. □

Theorem 187 (Feller Processes and Feller Semigroups) *A Markov process is a Feller process if and only if its evolution operators form a Feller semigroup.*

PROOF: Combine Lemmas 185 and 186. □

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 188 (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 163 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. 14.12 $K_t f(x)$ and $H_t f(x)$ are both right-continuous, their Laplace transforms are unique, so $K_t f(x) = H_t f(x)$. □

Theorem 189 (Feller Semigroups are Strongly Continuous) *Every Feller semigroup K_t with generator G is strongly continuous on* $\mathrm{Dom}(G)$.

PROOF: From Corollary 159, we have, as seen in Chapter 13, for all $t \geq 0$,

$$K_t f - f \quad = \quad \int_0^t K_s G f ds \tag{14.13}$$

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 190 (Cadlag Modifications Implied by a Kind of Modulus of Continuity) *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*

$$\inf_{\Gamma \in \mathcal{F}_s^X:\ \mathbb{P}(\Gamma)=1} \ \sup_{s,t \in T:\ s \leq t \leq s+\delta} \mathbb{P}\left(\omega:\ \rho(X(s,\omega), X(t,\omega)) \geq \epsilon,\ \omega \in \Gamma | \mathcal{F}_s^X\right) \tag{14.14}$$

If, for all ϵ*,*

$$\lim_{\delta \to 0} \alpha(\epsilon, \delta) = 0 \tag{14.15}$$

then X *has a cadlag version.*

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

Lemma 191 (Markov Processes Have Cadlag Versions When They Don't Move Too Fast (in Probability)) *Let* X *be a separable homogeneous Markov process. Define*

$$\alpha(\epsilon, \delta) = \sup_{t \in T:\ 0 \leq t \leq \delta;\ x \in \Xi} \mathbb{P}\left(\rho(X_x(t), x) \geq \epsilon\right) \tag{14.16}$$

If, for every $\epsilon > 0$*,*

$$\lim_{\delta \to 0} \alpha(\epsilon, \delta) = 0 \tag{14.17}$$

then X *has a cadlag version.*

Proof: The α in this lemma is clearly the α in the preceding proposition (190), using the fact that X is Markovian with respect to its natural filtration (Theorem 112) and homogeneous. □

Lemma 192 (Markov Processes Have Cadlag Versions If They Don't Move Too Fast (in Expectation)) *A separable homogeneous Markov process* X *has a cadlag version if*

$$\lim_{\delta \downarrow 0} \sup_{x \in \Xi,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(t), x)\right] = 0 \tag{14.18}$$

Proof: Start with the Markov inequality.

$$\forall x, t > 0, \epsilon > 0,\ \mathbb{P}\left(\rho(X_x(t), x) \geq \epsilon\right) \leq \frac{\mathbf{E}\left[\rho(X_x(t), x)\right]}{\epsilon} \tag{14.19}$$

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

$$\forall \delta > 0, \epsilon > 0,\ \sup_{x,\ 0 \leq t \leq \delta} \mathbb{P}\left(\rho(X_x(\delta), x) \geq \epsilon\right) \leq \frac{1}{\epsilon} \sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(\delta), x)\right] \tag{14.21}$$

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

$$\lim_{\delta \to 0} \alpha(\epsilon, \delta) \leq \frac{1}{\epsilon} \lim_{\delta \downarrow 0} \sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[\rho(X_x(\delta), x)\right] = 0 \tag{14.22}$$

So the preceding lemma (191) applies. □

Theorem 193 (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.[2] Since the Feller semigroup is strongly continuous on the domain of its generator (Theorem 189), and that domain is dense in C_0 by the Hille-Yosida Theorem (163), we can pick our f_k to be in this class. The strong continuity is with respect to the C_0 norm, so $\sup_x |K_t f(x) - K_s f(x)| = \sup_x |K_s(K_{t-s} f(x) - f(x))| \rightarrow 0$ as $t - s \rightarrow 0$, 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 $\sup_{x,\ 0 \leq t \leq \delta} \mathbf{E}\left[|f(X_x(t)) - f(x)|\right] \rightarrow 0$ as $\delta \rightarrow 0$. Now Lemma 192 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 194 (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,

$$\mathbf{E}\left[f(X(\tau + t)) | \mathcal{F}^{X+}_{\tau}\right] = K_t f(X(\tau)) \tag{14.23}$$

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

$$\tau_h \equiv \inf_u \{u \geq \tau :\ u = kh,\ k \in \mathbb{N}\} \tag{14.24}$$

Now τ_h is almost surely finite (because τ is), and $\tau_h \rightarrow \tau$ a.s. as $h \rightarrow 0$. 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 $m \in \mathbb{N}$,

$$\mathbf{E}\left[f(X_h(n+m)) | \mathcal{F}^X_n\right] = K_{mh} f(X_h(n)) \tag{14.25}$$

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

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

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

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

$$\mathbf{E}\left[f(X(\tau_h + t))\mathbf{1}_B\right] = \mathbf{E}\left[K_t f(X(\tau_h))\mathbf{1}_B\right] \quad (14.27)$$

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

where we let $h \downarrow 0$, and invoke the fact that $X(t)$ is right-continuous (Theorem 193) and $K_t f$ is continuous. Since this holds for arbitrary $B \in \mathcal{F}^{X+}_\tau$, and $K_t f(X(\tau))$ has to be $\mathcal{F}^{X+}_\tau$-measurable, we have that

$$\mathbf{E}\left[f(X(\tau + t))|\mathcal{F}^{X+}_\tau\right] = K_t f(X(\tau)) \quad (14.29)$$

as required. □

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

Theorem 195 (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 $f \in \mathrm{Dom}(G)$,*

$$\mathbf{E}\left[f(X(\beta)) - f(X(\alpha))\right] = \mathbf{E}\left[\int_\alpha^\beta Gf(X(t))dt\right] \quad (14.30)$$

Proof: Exercise 14.4. □

Remark: A large number of results very similar to Eq. 14.30 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!

14.3 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 $\mathbf{E}\left[K_\tau f(x)\right] = \lambda R_\lambda f(x)$.*

Exercise 14.2 (The First Pair of Feller Properties) *Prove Lemma 185. 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 186.*

Exercise 14.4 (Dynkin's Formula) *Prove Theorem 195.*

Exercise 14.5 (Lévy and Feller Processes) *Is every Lévy process a Feller process? If yes, prove it. If not, provide a counter-example, and try to find a sufficient condition for a Lévy process to be Feller.*