Chapter 15

Convergence of Feller Processes

This chapter looks at the convergence of sequences of Feller processes to a limiting process.

Section 15.1 lays some ground work concerning weak convergence of processes with cadlag sample paths.

Section 15.2 states and proves the central theorem about the convergence of sequences of Feller processes.

Section 15.3 examines a particularly important special case, the approximation of ordinary differential equations by pure-jump Markov processes.

15.1 Weak Convergence of Processes with Cadlag Paths (The Skorokhod Topology)

Recall that a sequence of random variables $X_1, X_2, \ldots$ converges in distribution on X, or weakly converges on X, $X_n \stackrel{d}{\rightarrow} X$, if and only if $\mathbf{E}[f(X_n)] \rightarrow \mathbf{E}[f(X)]$, for all bounded, continuous functions f. This is still true when X_n are random functions, i.e., stochastic processes, only now the relevant functions f are functionals of the sample paths.

Definition 161 (Convergence in Finite-Dimensional Distribution) *Random processes X_n on T* converge in finite-dimensional distribution *on X, $X_n \stackrel{fd}{\rightarrow} X$, when, $\forall J \in \mathrm{Fin}(T)$, $X_n(J) \stackrel{d}{\rightarrow} X(J)$.*

Proposition 162 *Convergence in finite-dimensional distribution is necessary but not sufficient for convergence in distribution.*

PROOF: Necessity is obvious: the coordinate projections π_t are continuous functionals of the sample path, so they must converge if the distributions converge. Insufficiency stems from the problem that, even if a sequence of X_n all have sample paths in some set U, the limiting process might not: recall our example (78) of the version of the Wiener process with unmeasurable suprema. □

Definition 163 (The Space D) *By $\mathbf{D}(T, \Xi)$ we denote the space of all cadlag functions from T to Ξ. By default, $\mathbf{D}$ will mean $\mathbf{D}(\mathbb{R}^+, \Xi)$.*

$\mathbf{D}$ admits of multiple topologies. For most purposes, the most convenient one is the *Skorokhod topology*, a.k.a. the *J_1 topology* or the *Skorokhod J_1 topology*, which makes $\mathbf{D}(\Xi)$ a complete separable metric space when Ξ is itself complete and separable. (See Appendix A2 of Kallenberg.) For our purposes, we need only the following notion and theorem.

Definition 164 (Modified Modulus of Continuity) *The* modified modulus of continuity *of a function $x \in \mathbf{D}(T, \Xi)$ at time $t \in T$ and scale $h > 0$ is given by*

$$w(x, t, h) \equiv \inf_{(I_k)} \max_k \sup_{r,s \in I_k} \rho(x(s), x(r)) \tag{15.1}$$

where the infimum is over partitions of $[0, t)$ into half-open intervals whose length is at least h (except possibly for the last one). Because x is cadlag, for fixed x and t, $w(x, t, h) \to 0$ as $h \to 0$.

Theorem 165 (Weak Convergence in $\mathbf{D}(\mathbb{R}^+, \Xi)$) *Let Ξ be a complete, separable metric space. Then a sequence of random functions $X_1, X_2, \ldots \in \mathbf{D}(\mathbb{R}^+, \Xi)$ converges in distribution to $X \in \mathbf{D}$ if and only if*

i The set $T_c = \{t \in T : X(t) = X(t^-)\}$ has a countable dense subset T_0, and the finite-dimensional distributions of the X_n converge on those of X on T_0.

ii For every t,

$$\lim_{h \to 0} \limsup_{n \to \infty} \mathbf{E}\left[w(X_n, t, h) \wedge 1\right] = 0 \tag{15.2}$$

PROOF: See Kallenberg, Theorem 16.10, pp. 313–314. □

Theorem 166 (Sufficient Condition for Weak Convergence) *The following three conditions are all equivalent, and all imply condition (ii) in Theorem 165.*

1. *For any sequence of a.s.-finite $\mathcal{F}^{X_n}$-optional times τ_n and positive constants $h_n \to 0$,*

$$\rho(X_n(\tau_n), X_n(\tau_n + h_n)) \xrightarrow{P} 0 \tag{15.3}$$

2. *For all $t > 0$, for all*

$$\lim_{h\rightarrow 0}\limsup_{n\rightarrow\infty}\sup_{\sigma,\tau}\mathbf{E}\left[\rho(X_n(\sigma),X_n(\tau))\wedge 1\right] = 0 \quad (15.4)$$

where σ and τ are $\mathcal{F}^{X_n}$-optional times $\sigma,\tau \leq t$, with $\sigma \leq \tau \leq \tau + h$.

3. *For all $t > 0$,*

$$\lim_{\delta\rightarrow 0}\limsup_{n\rightarrow\infty}\sup_{\tau\leq t}\sup_{0\leq h\leq\delta}\mathbf{E}\left[\rho(X_n(\tau),X_n(\tau+h))\wedge 1\right] = 0 \quad (15.5)$$

where the supremum in τ runs over all F^{X_n}-optional times $\leq t$.

PROOF: See Kallenberg, Theorem 16.11, pp. 314–315. □

15.2 Convergence of Feller Processes

We need some technical notions about generators.

Definition 167 (Closed and Closable Generators, Closures) *A linear operator O on a Banach space $\mathcal{B}$ is* closed *if its* graph *— $\left\{f,g\in\mathcal{B}^2 : f\in\mathrm{Dom}(O),\ g = Of\right\}$ — is a closed set. An operator is* closable *if the closure of its graph is a function (and not just a relation). The* closure *of a closable operator is that function.*

Notice, by the way, that because O is linear, it is closable iff $f_n \rightarrow 0$ and $Af_n \rightarrow g$ implies $g = 0$.

Definition 168 (Core of an Operator) *Let O be a closed linear operator on a Banach space $\mathcal{B}$. A linear subspace $D \subseteq \mathrm{Dom}(O)$ is a* core *of O if the closure of O restricted to D is, again O.*

The idea of a core is that we can get away with knowing how the operator works on a linear subspace, which is often much easier to deal with, rather than controlling how it acts on its whole domain.

Proposition 169 *The generator of every Feller semigroup is closed.*

PROOF: We need to show that the graph of G contains all of its limit points, that is, if $f_n \in \mathrm{Dom}(G)$ converges (in L_∞) on f, and $Gf_n \rightarrow g$, then $f \in \mathrm{Dom}(G)$ and $Gf = g$. First we show that $f \in \mathrm{Dom}(G)$.

$$\lim_{n\rightarrow\infty}(I-G)f_n = \lim_n f_n - \lim_n Gf_n \quad (15.6)$$

$$= f - g \quad (15.7)$$

But $(I - G)^{-1} = R_1$. Since this is a bounded linear operator, we can exchange applying the inverse and taking the limit, i.e.,

$$R_1 \lim_n (I - G) f_n = R_1(f - g) \tag{15.8}$$
$$\lim_n R_1 (I - G) f_n = R_1(f - g) \tag{15.9}$$
$$\lim_n f_n = R_1(f - g) \tag{15.10}$$
$$f = R_1(f - g) \tag{15.11}$$

Since the range of the resolvents is contained in the domain of the generator, $f \in \mathrm{Dom}(G)$. We can therefore say that $f - g = (I - G)f$, which implies that $Gf = g$. Hence, the graph of G contains all its limit points, and G is closed. □

Theorem 170 *Let X_n be a sequence of Feller processes with semigroups $K_{n,t}$ and generators G_n, and X be another Feller process with semigroup K_t and a generator G containing a core D. Then the following are equivalent.*

1. *If $f \in D$, there exists a sequence of $f_n \in \mathrm{Dom}(G_n)$ such that $\|f_n - f\|_\infty \to 0$ and $\|A_n f_n - Af\|_\infty \to 0$.*
2. *$K_{n,t} \to K_t$ for every $t > 0$*
3. *$\|K_{n,t} f - K_t f\|_\infty \to 0$ uniformly over $f \in C_0$ for bounded positive t*
4. *If $X_n(0) \xrightarrow{d} X(0)$ in Ξ, then $X_n \xrightarrow{d} X$ in $\mathbf{D}$.*

Proof: See Kallenberg, Theorem 19.25, p. 385. □

Remark. The important versions of the property above are the second — convergence of the semigroups — and the fourth — converge in distribution of the processes. The other two are there to simplify the proof.

15.3 Approximation of Ordinary Differential Equations by Markov Processes

The following result, due to Following Kurtz (1970, 1971), is essentially an application of Theorem 170.

First, recall that continuous-time, discrete-state Markov processes work essentially like a combination of a Poisson process (giving the time of transitions) with a Markov chain (giving the state moved to on transitions). This can be generalized to continuous-time, continuous-state processes, of what are called "pure jump" type.

Definition 171 (Pure Jump Markov Process) *A continuous-parameter Markov process is a* pure jump *process when its sample paths are piece-wise constant. For each state, there is an exponential distribution of times spent in that state, whose parameter is denoted $\lambda(x)$, and a transition probability kernel or exit distribution $\mu(x, B)$.*

Observe that pure-jump Markov processes always have cadlag sample paths. Also observe that the average amount of time the process spends in state x, once it jumps there, is $1/\lambda(x)$. So the time-average "velocity", i.e., rate of change, starting from x,

$$\lambda(x)\int_{\Xi}(y-x)\mu(x,dy)$$

Theorem 172 *Let X_n be a sequence of pure-jump Markov processes with state spaces Ξ_n, holding time parameters λ_n and transition probabilities μ_n. Suppose that, for all n Ξ_n is a Borel-measurable subset of $\mathbb{R}^k$ for some k. Let Ξ be another measurable subset of $\mathbb{R}^k$, on which there exists a function $F(x)$ such that $|F(x)-F(y)| \leq M|x-y|$ for some constant M. Suppose all of the following conditions holds.*

1. *The time-averaged rate of change is always finite:*

$$\sup_n \sup_{x\in\Xi_n\cap\Xi} \lambda_n(x)\int_{\Xi_n}|y-x|\mu_n(x,dy) \quad < \quad \infty \tag{15.12}$$

2. *There exists a positive sequence $\epsilon_n \rightarrow 0$ such that*

$$\lim_{n\rightarrow\infty} \sup_{x\in\Xi_n\cap\Xi} \lambda_n(x)\int_{|y-x|>\epsilon}|y-x|\mu_n(x,dy) \quad = \quad 0 \tag{15.13}$$

3. *The worst-case difference between $F(x)$ and the time-averaged rates of change goes to zero:*

$$\lim_{n\rightarrow\infty} \sup_{x\in\Xi_n\cap\Xi} \left|F(x)-\lambda_n(x)\int(y-x)\mu_n(x,dy)\right| \quad = \quad 0 \tag{15.14}$$

Let $X(s,x_0)$ be the solution to the initial-value problem where the differential is given by F, i.e., for each $0 \leq s \leq t$,

$$\frac{\partial}{\partial s}X(s,x_0) \quad = \quad F(X(s,x_0)) \tag{15.15}$$

$$X(0,x_0) \quad = \quad x_0 \tag{15.16}$$

and suppose there exists an $\eta > 0$ such that, for all n,

$$\Xi_n \cap \left\{y\in\mathbb{R}^k : \inf_{0\leq s\leq t}|y-X(s,x_0)| \leq \eta\right\} \quad \subseteq \quad \Xi \tag{15.17}$$

Then $\lim X_n(0) = x_0$ implies that, for every $\delta > 0$,

$$\lim_{n\rightarrow\infty} \mathbb{P}\left(\sup_{0\leq s\leq t}|X_n(s)-X(s,x_0)| > \delta\right) \quad = \quad 0 \tag{15.18}$$

The first conditions on the X_n basically make sure that they are Feller processes. The subsequent ones make sure that the mean time-averaged rate of change of the jump processes converges on the instantaneous rate of change of the differential equation, and that, if we're sufficiently close to the solution of the differential equation in $\mathbb{R}^k$, we're not in some weird way outside the relevant domains of definition. Even though Theorem 170 is about weak convergence, converging in distribution on a non-random object is the same as converging in probability, which is how we get uniform-in-time convergence in probability for a conclusion.

There are, broadly speaking, two kinds of uses for this result. One kind is practical, and has to do with justifying convenient approximations. If n is large, we can get away with using an ODE instead of the noisy stochastic scheme, or alternately we can use stochastic simulation to approximate the solutions of ugly ODEs. The other kind is theoretical, about showing that the large-population limit behaves deterministically, even when the individual behavior is stochastic *and strongly dependent over time.*