Chapter 9

Markov Processes

This lecture begins our study of Markov processes.

Section 9.1 is mainly "ideological": it formally defines the Markov property for one-parameter processes, and explains why it is a natural generalization of both complete determinism and complete statistical independence.

Section 9.2 introduces the description of Markov processes in terms of their transition probabilities and proves the existence of such processes.

9.1 The Correct Line on the Markov Property

The Markov property is the independence of the future from the past, given the present. Let us be more formal.

Definition 99 (Markov Property) *A one-parameter process X is a Markov process with respect to a filtration $\mathcal{F}$ when X_t is adapted to the filtration, and, for any $s > t$, X_s is independent of $\mathcal{F}_t$ given X_t, $X_s \perp\!\!\!\perp \mathcal{F}_t | X_t$. If no filtration is mentioned, it may be assumed to be the natural one generated by X. If X is also conditionally stationary, then it is a* time-homogeneous *(or just* homogeneous*) Markov process.*

Lemma 100 *Let X_t^+ stand for the collection of X_u, $u > t$. If X is Markov, then $X_t^+ \perp\!\!\!\perp \mathcal{F}_t | X_t$.*

PROOF: See Exercise 9.1. □

There are two routes to the Markov property. One is the path followed by Markov himself, of desiring to weaken the assumption of strict statistical independence between variables to mere conditional independence. In fact, Markov specifically wanted to show that independence was *not* a necessary condition for the law of large numbers to hold, because his arch-enemy claimed that it was, and used that as grounds for believing in free will and Christianity.[1] It turns out that all the key limit theorems of probability — the weak and strong laws of large numbers, the central limit theorem, etc. — work perfectly well for Markov processes, as well as for IID variables.

The other route to the Markov property begins with completely deterministic systems in physics and dynamics. The *state* of a deterministic dynamical system is some variable which fixes the value of all present and future observables. As a consequence, the present state determines the state at all future times. However, strictly deterministic systems are rather thin on the ground, so a natural generalization is to say that the present state determines the *distribution* of future states. This is precisely the Markov property.

Remarkably enough, it is possible to represent any one-parameter stochastic process X as a noisy function of a Markov process Z. The shift operators give a trivial way of doing this, where the Z process is not just homogeneous but actually fully deterministic. An equally trivial, but slightly more probabilistic, approach is to set $Z_t = X_t^-$, the complete past up to and including time t. (This is not necessarily homogeneous.) It turns out that, subject to mild topological conditions on the space X lives in, there is a unique *non-trivial* representation where $Z_t = \epsilon(X_t^-)$ for some function ϵ, Z_t is a homogeneous Markov process, and $X_u \perp\!\!\!\perp \sigma(\{X_t, t \leq u\})|Z_t$. (See Knight (1975, 1992).) We may explore such *predictive Markovian representations* at the end of the course, if time permits.

9.2 Transition Probability Kernels

The most obvious way to specify a Markov process is to say what its transition probabilities are. That is, we want to know $\mathbb{P}(X_s \in B | X_t = x)$ for every $s > t$, $x \in \Xi$, and $B \in \mathcal{X}$. Probability kernels (Definition 30) were invented to let us do just this.

Definition 101 (Product of Probability Kernels) *Let μ and ν be two probability kernels from Ξ to Ξ. Then their product $\mu\nu$ is a kernel from Ξ to Ξ, defined by*

$$(\mu\nu)(x, B) \equiv \int \mu(x, dy)\nu(y, B) \quad (9.1)$$

$$= (\mu \otimes \nu)(x, \Xi \times B) \quad (9.2)$$

Intuitively, all the product does is say that the probability of starting at the point x and landing in the set B is equal the probability of first going to y and then ending in B, integrated over all intermediate points y. (Strictly speaking, there is an abuse of notation in Eq. 9.2, since the second kernel in a composition $\otimes$ should be defined over a product space, here $\Xi \times \Xi$. So suppose we have such a

[1] I am not making this up. See Basharin *et al.* (2004) for a nice discussion of the origin of Markov chains and of Markov's original, highly elegant, work on them. There is a translation of Markov's original paper in an appendix to Howard (1971), and I dare say other places as well.

kernel ν', only $\nu'((x,y),B) = \nu(y,B)$.) Finally, observe that if $\mu(x,\cdot) = \delta_x$, the delta function at x, then $(\mu\nu)(x,B) = \nu(x,B)$, and similarly that $(\nu\mu)(x,B) = \nu(x,B)$.

Definition 102 *For every $(t,s) \in T \times T$, $s \geq t$, let $\mu_{t,s}$ be a probability kernel from Ξ to Ξ. These probability kernels form a* transition semi-group *when*

1. *For all t, $\mu_{t,t}(x,\cdot) = \delta_x$.*
2. *For any $t \leq s \leq u \in T$, $\mu_{t,u} = \mu_{t,s}\mu_{s,u}$.*

A transition semi-group for which $\forall t \leq s \in T$, $\mu_{t,s} = \mu_{0,s-t} \equiv \mu_{s-t}$ is homogeneous.

As with the shift semi-group, this is really a monoid (because $\mu_{t,t}$ acts as the identity).

The major theorem is the existence of Markov processes with specified transition kernels.

Theorem 103 *Let $\mu_{t,s}$ be a transition semi-group and ν_t a collection of distributions on a Borel space Ξ. If*

$$\nu_s = \nu_t \mu_{t,s} \tag{9.3}$$

then there exists a Markov process X such that

$$\forall t,\ \mathcal{L}(X_t) = \nu_t \tag{9.4}$$

$$\forall t_1 \leq t_2 \leq \ldots \leq t_n,\ \mathcal{L}(X_{t_1}, X_{t_2} \ldots X_{t_n}) = \nu_{t_1} \otimes \mu_{t_1,t_2} \otimes \ldots \otimes \mu_{t_{n-1},t_n} \tag{9.5}$$

Conversely, if X is a Markov process with values in Ξ, then there exist distributions ν_t and a transition kernel semi-group $\mu_{t,s}$ such that Equations 9.4 and 9.3 hold, and

$$\mathbb{P}(X_s \in B|\mathcal{F}_t) = \mu_{t,s} \text{ a.s.} \tag{9.6}$$

PROOF: (*From transition kernels to a Markov process.*) For any finite set of times $J = \{t_1, \ldots t_n\}$ (in ascending order), define a distribution on Ξ_J as

$$\nu_J \equiv \nu_{t_1} \otimes \mu_{t_1,t_2} \otimes \ldots \otimes \mu_{t_{n-1},t_n} \tag{9.7}$$

It is easily checked, using point (2) in the definition of a transition kernel semi-group (Definition 102), that the ν_J form a projective family of distributions. Thus, by the Kolmogorov Extension Theorem (Theorem 29), there exists a stochastic process whose finite-dimensional distributions are the ν_J. Now pick a J of size n, and two sets, $B \in \mathcal{X}^{n-1}$ and $C \in \mathcal{X}$.

$$\begin{aligned} \mathbb{P}(X_J \in B \times C) &= \nu_J(B \times C) && (9.8)\\ &= \mathbf{E}[\mathbf{1}_{B\times C}(X_J)] && (9.9)\\ &= \mathbf{E}\left[\mathbf{1}_B(X_{J\setminus t_n})\mu_{t_{n-1},t_n}(X_{t_{n-1}}, C)\right] && (9.10) \end{aligned}$$

Set $\mathcal{F}_t$ to be the natural filtration, $\sigma(\{X_u, u \leq s\})$. If $A \in \mathcal{F}_s$ for some $s \leq t$, then by the usual generating class arguments we have

$$\mathbb{P}\left(X_t \in C, X_s^- \in A\right) = \mathbf{E}\left[\mathbf{1}_A \mu_{s,t}(X_s, C)\right] \quad (9.11)$$

$$\mathbb{P}\left(X_t \in C | \mathcal{F}_s\right) = \mu_{s,t}(X_s, C) \quad (9.12)$$

i.e., $X_t \perp\!\!\!\perp \mathcal{F}_s | X_s$, as was to be shown.

(*From the Markov property to the transition kernels.*) From the Markov property, for any measurable set $C \in \mathcal{X}$, $\mathbb{P}\left(X_t \in C | \mathcal{F}_s\right)$ is a function of X_s alone. So define the kernel $\mu_{s,t}$ by $\mu_{s,t}(x, C) = \mathbb{P}\left(X_t \in C | X_s = x\right)$, with a possible measure-0 exceptional set from (ultimately) the Radon-Nikodym theorem. (The fact that Ξ is Borel guarantees the existence of a regular version of this conditional probability.) We get the semi-group property for these kernels thus: pick any three times $t \leq s \leq u$, and a measurable set $C \subseteq \Xi$. Then

$$\mu_{t,u}(X_t, C) = \mathbb{P}\left(X_u \in C | \mathcal{F}_t\right) \quad (9.13)$$

$$= \mathbb{P}\left(X_u \in C, X_s \in \Xi | \mathcal{F}_t\right) \quad (9.14)$$

$$= (\mu_{t,s} \otimes \mu_{s,u})(X_t, \Xi \times C) \quad (9.15)$$

$$= (\mu_{t,s}\mu_{s,u})(X_t, C) \quad (9.16)$$

The argument to get Eq. 9.3 is similar. □

Note: For one-sided discrete-parameter processes, we could use the Ionescu-Tulcea Extension Theorem 33 to go from a transition kernel semi-group to a Markov process, even if Ξ is not a Borel space.

Definition 104 *Let X be a homogeneous Markov process with transition kernels μ_t. A distribution ν on Ξ is* invariant *when, $\forall t$, $\nu = \nu\mu_t$, i.e.,*

$$(\nu\mu_t)(B) \equiv \int \nu(dx)\mu_t(x, B) \quad (9.17)$$

$$= \nu(B) \quad (9.18)$$

ν is also called an equilibrium distribution.

The term "equilibrium" comes from statistical physics, where however its meaning is a bit more strict, in that "detailed balance" must also be satisified: for any two sets $A, B \in \mathcal{X}$,

$$\int \nu(dx)\mathbf{1}_A \mu_t(x, B) = \int \nu(dx)\mathbf{1}_B \mu_t(x, A) \quad (9.19)$$

i.e., the flow of probability from A to B must equal the flow in the opposite direction. Much confusion has resulted from neglecting the distinction between equilibrium in the strict sense of detailed balance and equilibrium in the weaker sense of invariance.

Theorem 105 *Suppose X is homogeneous, and $\mathcal{L}(X_t) = \nu$, where ν is an invariant distribution. Then the process X_t^+ is stationary.*

PROOF: Exercise 9.4. □

9.3 Exercises

Exercise 9.1 *Prove Lemma 100.*

Exercise 9.2 *Show that if X is a Markov process, then, for any $t \in T$, X_t^+ is a one-sided Markov process.*

Exercise 9.3 *Let X be a continuous-parameter Markov process, and t_n a countable set of strictly increasing indices. Set $Y_n = X_{t_n}$. Is Y_n a Markov process? If X is homogeneous, is Y also homogeneous? Does either answer change if $t_n = nt$ for some constant interval $t > 0$?*

Exercise 9.4 *Prove Theorem 105.*