Chapter 10

Alternate Characterizations of Markov Processes

This lecture introduces two ways of characterizing Markov processes other than through their transition probabilities.

Section 10.1 addresses a question raised in the last class, about when being Markovian relative to one filtration implies being Markov relative to another.

Section 10.2 describes discrete-parameter Markov processes as transformations of sequences of IID uniform variables.

Section 10.3 describes Markov processes in terms of measure-preserving transformations (Markov operators), and shows this is equivalent to the transition-probability view.

10.1 The Markov Property Under Multiple Filtrations

In the last lecture, we defined what it is for a process to be Markovian relative to a given filtration $\mathcal{F}_t$. The question came up in class of when knowing that X Markov with respect to one filtration $\mathcal{F}_t$ will allow us to deduce that it is Markov with respect to another, say $\mathcal{G}_t$.

To begin with, let's introduce a little notation.

Definition 106 (Natural Filtration) *The natural filtration for a stochastic process X is $\mathcal{F}_t^X \equiv \sigma(\{X_u,\ u \leq t\})$. Obviously, every process X is adapted to $\mathcal{F}_t^X$.*

Definition 107 (Comparison of Filtrations) *A filtration $\mathcal{G}_t$ is* finer than *or* more refined than *or* a refinement of *$\mathcal{F}_t$, $\mathcal{F}_t \prec \mathcal{G}_t$, if, for all t, $\mathcal{F}_t \subseteq G_t$, and at least sometimes the inequality is strict. $\mathcal{F}_t$ is* coarser *or* less fine *than $\mathcal{G}_t$. If $\mathcal{F}_t \prec \mathcal{G}_t$ or $\mathcal{F}_t = \mathcal{G}_t$, we write $\mathcal{F}_t \preceq \mathcal{G}_t$.*

Lemma 108 *If X is adapted to $\mathcal{G}_t$, then $\mathcal{F}_t^X \preceq \mathcal{G}_t$.*

PROOF: For each t, X_t is $\mathcal{G}_t$ measurable. But $\mathcal{F}_t^X$ is, by construction, the smallest σ-algebra with respect to which X_t is measurable, so, for every t, $\mathcal{F}_t^X \subseteq \mathcal{G}_t$, and the result follows. □

Theorem 109 *If X is Markovian with respect to $\mathcal{G}_t$, then it is Markovian with respect to any coarser filtration to which it is adapted, and in particular with respect to its natural filtration.*

PROOF: Use the smoothing property of conditional expectations: For any two σ-fields $\mathcal{F} \subset \mathcal{G}$ and random variable Y, $\mathbf{E}[Y|\mathcal{F}] = \mathbf{E}[\mathbf{E}[Y|\mathcal{G}]|\mathcal{F}]$ a.s. So, if $\mathcal{F}_t$ is coarser than $\mathcal{G}_t$, and X is Markovian with respect to the latter, for any function $f \in L_1$ and time $s > t$,

$$\begin{aligned}
\mathbf{E}[f(X_s)|\mathcal{F}_t] &= \mathbf{E}[\mathbf{E}[f(X_s)|\mathcal{G}_t]|\mathcal{F}_t] \ a.s. && (10.1)\\
&= \mathbf{E}[\mathbf{E}[f(X_s)|X_t]|\mathcal{F}_t] && (10.2)\\
&= \mathbf{E}[f(X_s)|X_t] && (10.3)
\end{aligned}$$

where the last line uses the facts that (i) $\mathbf{E}[f(X_s)|X_t]$ is a function X_t, (ii) X is adapted to $\mathcal{F}_t$, so X_t is $\mathcal{F}_t$-measurable, and (iii) if Y is $\mathcal{F}$-measurable, then $\mathbf{E}[Y|\mathcal{F}] = Y$. Since this holds for all $f \in L_1$, it holds in particular for $\mathbf{1}_A$, where A is any measurable set, and this established the conditional independence which constitutes the Markov property. Since (Lemma 108) the natural filtration is the coarsest filtration to which X is adapted, the remainder of the theorem follows. □

The converse is false, as the following example shows.

Example 110 *We revert to the symbolic dynamics of the logistic map, Examples 39 and 40. Let S_1 be distributed on the unit interval with density $1/\pi\sqrt{s(1-s)}$, and let $S_n = 4S_{n-1}(1-S_{n-1})$. Finally, let $X_n = \mathbf{1}_{[0.5,1.0]}(S_n)$. It can be shown that the X_n are a Markov process with respect to their natural filtration; in fact, with respect to that filtration, they are independent and identically distributed Bernoulli variables with probability of success $1/2$. However, $\mathbb{P}\left(X_{n+1}|\mathcal{F}_n^S, X_n\right) \neq \mathbb{P}\left(X_{n+1}|X_n\right)$, since X_{n+1} is a* deterministic *function of S_n. But, clearly, $\mathcal{F}_n^S$ is a refinement of $\mathcal{F}_n^X$.*

The issue can be illustrated with graphical models (Spirtes *et al.*, 2001; Pearl, 1988). A discrete-time Markov process looks like Figure 10.1*a*. X_n blocks all the pasts from the past to the future (in the diagram, from left to right), so it produces the desired conditional independence. Now let's add another variable which actually drives the X_n (Figure 10.1*b*). If we can't measure the S_n variables, just the X_n ones, then it can still be the case that we've got the conditional independence among what we can see. But if we can see X_n as well as S_n — which is what refining the filtration amounts to — then simply conditioning on X_n does not block all the paths from the past of X to its future, and, generally speaking, we will lose the Markov property. Note that knowing

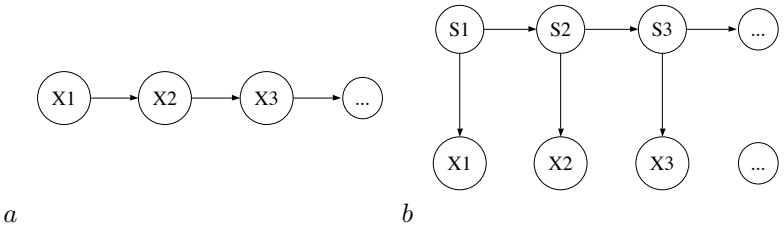

Figure 10.1: (a) Graphical model for a Markov chain. (b) Refining the filtration, say by conditioning on an additional random variable, can lead to a failure of the Markov property.

S_n *does* block all paths from past to future — so this remains a *hidden* Markov model. Markovian representation theory is about finding conditions under which we can get things to look like Figure 10.1b, even if we can't get them to look like Figure 10.1a.

10.2 Markov Sequences as Transduced Noise

A key theorem says that discrete-time Markov processes can be viewed as the result of applying a certain kind of filter to pure noise.

Theorem 111 *Let X be a one-sided discrete-parameter process taking values in a Borel space Ξ. X is Markov iff there are measurable functions $f_n : \Xi \times [0,1] \mapsto \Xi$ such that, for IID random variables $Z_n \sim U(0,1)$, all independent of X_1, $X_{n+1} = f_n(X_n, Z_n)$ almost surely. X is homogeneous iff $f_n = f$ for all n.*

Proof: Kallenberg, Proposition 8.6, p. 145. Notice that, in order to get the "only if" direction to work, Kallenberg invokes what we have as 26, which is where the assumptions that Ξ is a Borel space comes in. You should verify that the "if" direction does not require this assumption. □

Let us stick to the homogeneous case, and consider the function f in somewhat more detail.

In engineering or computer science, a *transducer* is an apparatus — really, a function — which takes a stream of inputs of one kind and produces a stream of outputs of another kind.

Definition 112 (Transducer) *A (deterministic) transducer is a sextuple $\langle \Sigma, \Upsilon, \Xi, f, h, s_0 \rangle$ where Σ, Υ and Ξ are, respectively, the state, input and output spaces, $f : \Sigma \times \Xi \mapsto \Sigma$ is the* state update function *or* state transition function, *$h : \Sigma \times \Upsilon \mapsto \Xi$ is the* measurement *or* observation *function, and $s_0 \in \Sigma$ is the* starting state. *(We shall assume both f and h are always measurable.) If h does not depend on its state argument, the transducer is* memoryless. *If f does not depend on its state argument, the transducer is* without after-effect.

It should be clear that if a memoryless transducer is presented with IID inputs, its output will be IID as well. What Theorem 111 says is that, if we have a transducer with memory (so that h depends on the state) but is without after-effect (so that f does not depend on the state), IID inputs will produce Markovian outputs, and conversely any reasonable Markov process can be represented in this way. Notice that if a transducer is without memory, we can replace it with an equivalent with a single state, and if it is without after-effect, we can identify Σ and Ξ.

Notice also that the two functions f and h determine a transition function where we use the input to update the state: $g : \Sigma \times \Upsilon \mapsto \Sigma$, where $g(s, y) = f(s, h(s, y))$. Thus, if the inputs are IID and uniformly distributed, then (Theorem 111) the successive states of the transducer are always Markovian. The question of which processes can be produced by noise-driven transducers is this intimately bound up with the question of Markovian representations. While, as mentioned, quite general stochastic processes *can* be put in this form (Knight, 1975, 1992), it is not necessarily possible to do this with a finite internal state space Σ, even when Ξ is finite. The distinction between finite and infinite Σ is crucial to theoretical computer science, and we might come back to it later, but

Two issues suggest themselves in connection with this material. One is whether, given a *two*-sided process, we can pull the same trick, and represent a Markovian X as a transformation of an IID sequence extending into the infinite past. (Remember that the theorem is for one-sided processes, and starts with an initial X_1.) This is more subtle than it seems at first glance, or even than it seemed to Norbert Wiener when he first posed the question (Wiener, 1958); for a detailed discussion, see Rosenblatt (1971), and, for recent set of applications, Wu (2005). The other question is whether the same trick can be pulled in continuous time; here much less is known.

10.3 Time-Evolution (Markov) Operators

Let's look again at the evolution of the one-dimensional distributions for a Markov process:

$$\nu_s = \nu_t \mu_{t,s} \tag{10.4}$$

$$\nu_s(B) = \int \nu_t(dx) \mu_{t,s}(x, B) \tag{10.5}$$

The transition kernels define linear operators taking distributions on Ξ to distributions on Ξ. This can be abstracted.

Definition 113 (Markov Operator) *Take any measure space $\Xi, \mathcal{X}, \mu$, and let L_1 be as usual the class of all μ-integrable generalized functions on Ξ. A linear operator $P : L_1 \mapsto L_1$ is a* Markov operator *when:*

1. *If $f \geq 0$ (a.e. μ), $Pf \geq 0$ (a.e. μ).*

2. *If $f \geq 0$ (a.e. μ), $\|Pf\| = \|f\|$.*
3. *$P\mathbf{1}_\Xi = \mathbf{1}_\Xi$.*
4. *If $f_n \downarrow 0$, then $Pf_n \downarrow 0$.*

Lemma 114 *Every probability kernel κ from Ξ to Ξ induces a Markov operator K,*

$$Kf(x) = \int \kappa(x, dy) f(y) \tag{10.6}$$

and conversely every operator defines a transition probability kernel,

$$\kappa(x, B) = K\mathbf{1}_B(x) \tag{10.7}$$

PROOF: Exercise 10.1. □

Clearly, if κ is part of a transition kernel semi-group, then the collection of induced Markov operators also forms a semi-group.

Theorem 115 (Markov operator semi-groups and Markov processes) *Let X be a Markov process with transition kernels $\mu_{t,s}$, and let $K_{t,s}$ be the corresponding semi-group of operators. Then for any $f \in L_1$,*

$$\mathbf{E}\left[f(X_s)|\mathcal{F}_t\right] = (K_{t,s}f)(X_t) \tag{10.8}$$

Conversely, let X be any stochastic process, and let $K_{t,s}$ be a semi-group of Markov operators such that Equation 10.8 is valid (a.s.). Then X is a Markov process.

PROOF: Exercise 10.2. □

Remark. The proof works because the expectations of all L_1 functions together determine a probability measure. If we knew of another collection of functions which also sufficed to determine a measure, then linear operators on that collection would work just as well, in the theorem, as do the Markov operators, which by definition apply to all of L_1. In particular, it is sometimes possible to define operators only on much smaller, more restricted collections of functions, which can have technical advantages. See Ethier and Kurtz (1986, ch. 4, sec. 1) for details.

The next two lemmas will prove useful in establishing asymptotic results.

Lemma 116 (Markov Operators are Contractions) *For any Markov operator P and $f \in L_1$,*

$$\|Pf\| \leq \|f\| \tag{10.9}$$

PROOF (after Lasota and Mackey (1994, prop. 3.1.1, pp. 38–39)): First, notice that $(Pf(x))^+ \leq Pf^+(x)$, because $(Pf(x))^+ = (Pf^+ - Pf^-)^+ = \max(0, Pf^+ - Pf^-) \leq \max(0, Pf^+) = Pf^+$. Similarly $(Pf(x))^- \leq Pf^-(x)$. Therefore $|Pf| \leq P|f|$, and then the statement follows by integration. □

Lemma 117 *For any Markov operator, and any $f, g \in L_1$, $\|P^n f - P^n g\|$ is non-increasing.*

PROOF: By linearity, $\|P^n f - P^n g\| = \|P^n(f-g)\|$. By the definition of P^n, $\|P^n(f-g)\| = \|PP^{n-1}(f-g)\|$. By the contraction property (Lemma 116), $\|PP^{n-1}(f-g)\| \leq \|P^{n-1}(f-g)\| = \|P^{n-1}f - P^{n-1}g\|$ (by linearity again). □

Theorem 118 *A probability measure ν is invariant for a homogeneous Markov process iff it is a fixed point of all the transition operators, $\nu K_t = \nu$.*

PROOF: Clear from the definitions! □

10.4 Exercises

Exercise 10.1 *Prove Lemma 114. Hint: you will want to use the fact that $\mathbf{1}_B \in L_1$ for all measurable sets B.*

Exercise 10.2 *Prove Theorem 115. Hint: in showing that a collection of operators determines a Markov process, try using mathematical induction on the finite-dimensional distributions.*