Markov chains, shift spaces, and the ergodic theorems.

In this section we will discuss some Markov chains and consequences of the ergodic theorems in this context. Markov chains are (in the discrete case) sequences of random variables attaining certain states (i.e., the values) at certain times (i.e., the index of the sequence). We denote the state space by S and assume that S is finite (and, therefore w.l.o.g., that $S \subseteq \mathbb{R}$). The theory we develop rather directly extends to countably infinite spaces, but we avoid this for (negligible) topological reasons. The extension to, say, $\mathbb{R}$-valued states is a bit trickier and to non-discrete times is quite involved. References include Norris – Markov Chains and Breiman – Probability

Definition

Let $X = S^{\mathbb{N}_0}$, $\mathcal{B} = \mathcal{B}(X)$. Let $\pi_n : X \longrightarrow S$, $(x_\ell)_{\ell\in\mathbb{N}_0} \mapsto x_n$ for $n \in \mathbb{N}_0$ the canonical projection.

We let $\mathcal{B}_n := \pi_n^{-1}(2^S) \subseteq \mathcal{B}$ and define the "natural filtration" $(\mathcal{F}_n)_{n\in\mathbb{N}_0}$ by

$$\mathcal{F}_n := \sigma\Big(\bigcup_{k=0}^{n} \mathcal{B}_k\Big) = \sigma(\pi_0, \dots, \pi_n).$$

Definition ((Coordinate process) Markov chain)

Let $X = S^{\mathbb{N}_0}$, $\mathcal{B} = \mathcal{B}(X)$, and μ a Borel probability measure on $(X, \mathcal{B})$. Then $(X, \mathcal{B}, \mu)$ is a "Markov chain" if

$$\forall n \in \mathbb{N}_0\ \forall \varphi : S \longrightarrow \mathbb{C} \quad \mathbb{E}(\varphi \circ \pi_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(\varphi \circ \pi_{n+1} \mid \mathcal{B}_n) \quad \mu\text{-a.s.}$$

Interpretation: Note that $\mathcal{F}_n$ is finite, generated by the finite partition

$$\Big\{ \mathcal{C}(s_0, \dots, s_n) = \{(x_\ell)_{\ell\in\mathbb{N}_0} \in X : \forall 0 \le \ell \le n \ \ x_\ell = s_\ell\} : (s_0,\dots,s_n) \in S^{n+1} \Big\}$$

Similarly, $\mathcal{B}_n$ is generated by the finite partition

$$\Big\{ \Sigma_n(s) = \bigsqcup_{(s_0,\dots,s_{n-1})\in S^n} \mathcal{C}(s_0, \dots, s_{n-1}, s) : s \in S \Big\} \quad (n \in \mathbb{N})$$

$$\{ \Sigma_0(s) = \mathcal{C}(s) : s \in S \} \quad (n = 0)$$

Hence for $\varphi = \mathbb{1}_{\{t\}}$, $t \in S$, we have

$$\mathbb{E}(\varphi \circ \pi_{n+1} \mid \mathcal{F}_n) = \sum_{\underline{s}\in S^{n+1}} \frac{\mathbb{E}((\mathbb{1}_{\{t\}} \circ \pi_{n+1}) \mathbb{1}_{\mathcal{C}(\underline{s})})}{\mu(\mathcal{C}(\underline{s}))} \mathbb{1}_{\mathcal{C}(\underline{s})}$$

$$= \sum_{\underline{s}\in S^{n+1}} \frac{\mathbb{E}(\mathbb{1}_{\mathcal{C}(\underline{s},t)})}{\mu(\mathcal{C}(\underline{s}))} \mathbb{1}_{\mathcal{C}(\underline{s})}.$$

Therefore $\forall (x_\ell)_{\ell\in\mathbb{N}_0} = \underline{x} \in X$

$$\mathbb{E}(\varphi \circ \pi_{n+1} \mid \mathcal{F}_n)(\underline{x}) = \frac{\mu(\mathcal{C}(x_0, \dots, x_n, t))}{\mu(\mathcal{C}(x_0, \dots, x_n))} = \mu(\pi_{n+1} = t \mid \pi_0 = x_0, \dots, \pi_n = x_n)$$

Similarly

$$\mathbb{E}(\varphi \circ \pi_{n+1} \mid \mathcal{B}_n)(\underline{x}) = \frac{\mu(\Sigma_{n+1}(t) \cap \Sigma_n(x_n))}{\mu(\Sigma_n(x_n))} = \mu(\pi_{n+1} = t \mid \pi_n = x_n).$$

Thus, the Markov property says that the probability that the coordinate process at time $n+1$ attains the value t given the past $(\pi_0, \dots, \pi_n)$ only depends on the immediate past π_n.

Definition:

Let μ be a probability measure on $(S^{\mathbb{N}_0}, \mathcal{B}(S^{\mathbb{N}_0}))$ such that $(S^{\mathbb{N}_0}, \mathcal{B}(S^{\mathbb{N}_0}), \mu)$ is a Markov chain. Let $\underline{p} \in \mathbb{R}^S$ be a probability vector and $P = (p_{st})_{s,t\in S} \in \mathbb{R}^{S\times S}$ such that the rows are probability vectors. Then $(S^{\mathbb{N}_0}, \mathcal{B}(S^{\mathbb{N}_0}), \mu)$ is a $(\underline{p}, P)$-Markov chain if

(i) $\forall s \in S \quad \mu(\mathcal{C}(s)) = p_s$,

(ii) $\forall n \in \mathbb{N}_0\ \forall s, t \in S \quad \mu(\pi_{n+1} = t \mid \pi_n = s) = p_{st}$.

Theorem

Let $(S^{\mathbb{N}_0}, \mathcal{B}(S^{\mathbb{N}_0}), \mu)$ be a $(\underline{p}, P)$-Markov chain. For all $m, n \in \mathbb{N}_0$ and for all $s, t \in S$

(i) $\mu(\pi_n = s) = (\underline{p}P^n)_s$ ← in part.

(ii) $\mu(\pi_n = t \mid \pi_0 = s) = \mu(\pi_{n+m} = t \mid \pi_m = s) = (P^n)_{st}$

End of lecture 17