Plan Countable-State Markov Chains

1. Transition Probabilities
2. Irreducibility and Recurrence
3. Classification of States and Chains

Next Time: Limit Theorems, General-state analogues

Reading: G&S 6.4, 6.5, 6.6
Homework 4 on-line today, due next week; Comments: typos, G&S terminology, hw

Definition 1. Let $(\mathcal{S}, \mathcal{B})$ be a measurable space. Let $P(x, A)$ be a function on $\mathcal{S} \times \mathcal{B}$ such that

1. For each $A \in \mathcal{B}$, $x \mapsto P(x, A)$ is a measurable function on $\mathcal{S}$.
2. For each $x \in \mathcal{S}$, $A \mapsto P(x, A)$ is a probability measure on $\mathcal{B}$.

We call such a P a *transition probability kernel.*

For a given $x \in \mathcal{S}$, let Y be a random variable with distribution $P(x, \cdot)$. Then, for a function g on $\mathcal{S}$, we can also write

$$P(x, g) \equiv \mathsf{E} g(Y) \equiv \int g(y) P\left(x, [y, y + dy)\right) \equiv \int g(y) P(x, dy), \tag{1}$$

where by slight abuse of notation we use dy to represent an infinitesimal interval around y. This mimics the relation between probabilities P and expected values E.

Reminder 2. We can view a probability measure ν on $(\mathcal{S}, \mathcal{B})$ in two equivalent ways:

1. As a set function that maps $A \in \mathcal{B}$ to $0 \le \nu(A) \le 1$ such that $\nu(\cdot)$ obeys all the rules of probability ($\nu(\mathcal{S}) = 1$, $\nu(\emptyset) = 0$, $\nu(U_i A_i) = \sum_i \nu(A_i)$ when the A_i are disjoint.
2. As an operator that maps measurable functions on $\mathcal{S}$ g to $-\infty \le \nu(g) \le \infty$ such that $\nu(\cdot)$ obeys all the rules of expected values.

Thus, the probability and expected value operators, P and E, are actually the same object.

Recall 3. If $P(s, A)$ is a transition probability kernel and $\mathcal{S}$ is countable, then we can write

$$\mathsf{P}(s, A) = \sum_{s' \in A} \mathsf{P}\{\, X_n = s' \mid X_{n-1} = s \,\} = \sum_{s' \in A} P(s, \{s'\}) \tag{2}$$

for any $A \subset \mathcal{S}$. In this case, we define $P_{ss'} \equiv P(s, s') \equiv P(s, \{s'\})$ and call it the *transition probability matrix.* A transition probability matrix must satisfy $P_{ss'} \equiv P(s, s') \ge 0$ and $P(s, \mathcal{S}) = 1 = \sum_{s' \in \mathcal{S}} P(s, s') \equiv \sum_{s' \in \mathcal{S}} P_{ss'}$.

Definition 4. A time-homogeneous, countable-state *Markov chain* with state space $\mathcal{S}$ with initial distribution μ and transition probability kernel $P(\cdot, \cdot)$ is an $S^{\mathbb{Z}_\oplus}$-valued stochastic process $X = (X_n)_{n \ge 0}$ such that

$$\mathsf{P}_\mu\{X_0 = s_0, \ldots, X_n = s_n\} = \mu(s_0) P(s_0, \{s_1\}) \cdots P(s_{n-1}, \{s_n\}), \tag{3}$$

for all $n \ge 0$ and all $s_0, \ldots, s_n \in \mathcal{S}$. We write P_μ to denote the chain has initial distribution μ.

Definition 5. A time-homogeneous, general-state *Markov chain* with state space $\mathcal{S}$ with initial distribution μ and transition probability kernel $P(\cdot,\cdot)$ is an $S^{\mathbb{Z}_\oplus}$-valued stochastic process $X = (X_n)_{n\geq 0}$ such that

$$\mathsf{P}_\mu\{X_0 \text{ near } s_0, \ldots, X_n \text{ near } s_n\} = \mu(s_0)P(s_0, ds_1)\cdots P(s_{n-1}, ds_n), \tag{4}$$

for all $n \geq 0$ and all $s_0, \ldots, s_n \in \mathcal{S}$, where by slight abuse of notation we use ds_i to represent an infinitesimal interval around s_i. Again, we write P_μ to denote the chain has initial distribution μ. A more formal (but less clear) expression of (4) is

$$\mathsf{P}_\mu(X_0 \in A_0, X_1 \in A_1, \ldots, X_n \in A_n) = \int_{A_0}\cdots\int_{A_{n-1}} \mu(ds_0)P(s_0, ds_1)\cdots P(s_{n-2}, ds_{n-1})P(s_{n-1}, A_n). \tag{5}$$

Reminder 6. If X is a random function $\mathcal{T} \to \mathcal{S}$, then the *finite-dimensional distributions* of X are collectively the distributions of random vactors $(X_{t_1}, \ldots, X_{t_n})$ for any $n \geq 1$ and $t_1, \ldots, t_n \in \mathcal{T}$. The rigorous construction of a σ-field on $\mathcal{S}^{\mathcal{T}}$ yields a consistent distribution over this space that is characterized by these finite-dimensional distributions.

Definition 7. We thus have three forms of the *Markov Property.*

1. The first is given in (3) and (4) and indicates that the finite-dimensional distributions of the process are determined solely by the initial distribution and the transition probabilities.
2. A conditional form of the first: for any $n \geq 1$ and $s_0, \ldots, s_n \in \mathcal{S}$,

$$\mathsf{P}_\mu\{X_n = s_n \mid X_{n-1} = s_{n-1}, \ldots, X_0 = s_0\} = P(s_{n-1}, s_n) \tag{6}$$

   for countable state spaces and

$$\mathsf{P}_\mu\{X_n \text{ near } s_n \mid X_{n-1} \text{ near } s_{n-1}, \ldots, X_0 \text{ near } s_0\} = P(s_{n-1}, ds_n) \tag{7}$$

   for general state spaces.
3. Let h be a bounded and measurable function on $\mathcal{S}^{\mathbb{Z}_\oplus}$. Then, for any $n \geq 1$,

$$\mathsf{E}(h(X_n, X_{n+1}, \ldots) \mid X_n \text{ near } s, X_{n-1}, \ldots, X_0) = \mathsf{E}_s h(X_1, X_2, \ldots). \tag{8}$$

Definition 8. Given a Markov chain with transition probability kernel $P(x, A)$, we can define the *n-step transition probabilities* $P^n(x, A)$ by induction as follows. For a transition probability matrix:

$$P^n(s, s') = \sum_{u\in\mathcal{S}} P(s, u)P^{n-1}(u, s') \implies P^n = P \cdot P^{n-1} = (P)^n \text{ as matrices.} \tag{9}$$

For general transition kernels:

$$P^n(s, A) = \int_{\mathcal{S}} P(s, du)P^{n-1}(u, A). \tag{12}$$

Theorem 9. *The Chapman-Kolmogorov Equations.*
Let $n, m \geq 0$. For a transition probability matrix P, we have:

$$P^{n+m} = P^n \cdot P^m, \qquad \text{(that is, } P^{n+m}(s, s') = \sum_{u \in \mathcal{S}} P^n(s, u) P^m(u, s')\text{)}. \tag{11}$$

For general transition kernels:

$$P^{n+m}(s, A) = \int_{\mathcal{S}} P^n(s, du) P^m(u, A). \tag{12}$$

Examples 10.

1. Simple Random Walk
2. Random Walk on a Pentagon
3. Embedded MC for G/M/1 queue
   $\mathcal{S} = \mathbb{Z}_{\oplus}$. Exponential$\langle\lambda\rangle$ service time. G is the service time F.

$$P(j, j+1-k) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} \, dG(t), \quad k \leq j, \tag{13}$$

$$P(j, 0) = \int_0^\infty \sum_{k=j+1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} \, dG(t). \tag{14}$$

4. The Flip-Flop and the d-adic Spin
5. The Finite, Infinite, and Isolated Black Hole
6. Given a countable-state Markov chain with initial distribution μ and transition probability matrix P, what is the distribution of X_n?

Definition 11. Given two states $s, s' \in \mathcal{S}$, we say that s' is *accessible* from s if $P^n(s, s') > 0$ for some $n \geq 0$. We say that s and s' *communicate* if each is accessible from the other. Denote this by $s \leftrightarrow s'$.
Claim: $\leftrightarrow$ is an equivalence class.
Proof: Reflexivity, Symmetry, Transitivity.

Definition 12. If there is only one equivalence class of communicating states, the Markov Chain is said to be *irreducible*

Definition 13. A set $A \subset \mathcal{S}$ is said to be *absorbing* if $P(x, A) = 1$ for all $x \in A$.

Decomposition 14. If X is not irreducible, then we can write

$$\mathcal{S} = D \cup \bigcup_i C_i, \tag{15}$$

where the sets are disjoint and each C_i is an absorbing, communicating class.

Question 15. What can happen if the chain is in D?

Proposition 16. If $C \subset \mathcal{S}$ is an absorbing, communicating class for a Markov chain X, then there exists an irreducible Markov Chain X^C with state-space C and whose transition probability kernel is given by $P_C(x, A) \equiv P(x, A)$ for $x \in C$.

Definition 17. For any state $s \in \mathcal{S}$, define the *period* of s by

$$d(s) = \gcd\{n \geq 1 \text{ such that } P^n(s,s) > 0\}. \tag{16}$$

This implies that $P^n(s,s) = 0$ unless $n = md(s)$ for some $m \in \mathbb{Z}_+$.

Theorem 18. d is a class function with respect to $\leftrightarrow$; that is, d is constant on communicating classes.

Proof. Let s, s' be members of the same class. Then, there is an n and an m such that $P^n(s,s') > 0$ and $P^m(s',s) > 0$. By the Chapman-Kolmogorov equations,

$$P^{n+m}(s,s) \geq P^n(s,s')P^m(s',s) > 0, \tag{17}$$

so $n+m$ is a multiple of $d(s)$.

Suppose k is not a multiple of $d(s)$; then neither is $k+n+m$:

$$0 = P^{k+m+n}(s,s) \geq P^n(s,s')P^k(s',s')P^m(s',s). \tag{18}$$

Thus, $P^k(s',s') = 0$ which implies that $d(s') \geq d(s)$.

Reverse the roles of s and s' to get equality.

Definition 19. An irreducible Markov Chain is said to be *aperiodic* if $d(s) \equiv 1$ for $s \in \mathcal{S}$. It is called *strongly aperiodic* if $P(s,s) > 0$ for some $s \in \mathcal{S}$.

Question 20. Can you find an example where these two notions differ?

Theorem 21. Let X be an irreducible, countable-state Markov chain with common period d. Then, there are disjoint sets $U_1, \ldots, U_d \subset \mathcal{S}$ such that

$$\mathcal{S} = \bigcup_{k=1}^{d} U_k, \tag{19}$$

and

$$\mathsf{P}(x, U_{k+1}) = 1 \quad \text{for } x \in U_k, \qquad k = 0, \ldots, d-1(\pmod d). \tag{20}$$

The sets $U_1, \ldots, U_d$ are called *cyclic classes* of X because X cycles through them successively.

It follows that that the Markov Chain $X^d = (X_d, X_{2d}, X_{3d}, \ldots)$ has transition probabilities P^d and each U_i is an absorbing, irreducible, aperiodic class.

Further, for $k = 0, \ldots, d-1$, if $\mu(U_k) = 1$, X^d is an irreducible, aperiodic Markov chain.

Pause for Breath 22. What does all this mean for the analysis of Markov Chains? Examples, thoughts, concerns, and questions.

Useful Random Variables 23. Let $A \subset \mathcal{S}$. Define

$$T_A = \inf\{n \geq 1 \text{ such that } X_n \in A\} \tag{21}$$

$$S_A = \inf\{n \geq 0 \text{ such that } X_n \in A\} \tag{22}$$

$$O_A = \sum_{n=1}^{\infty} 1\{X_n \in A\}. \tag{23}$$

These are called, respectively, the *first return time*, the *first hitting time*, and the *occupation time* of A. As we will see, these random variables provide a great deal of information about the behavior of the chain.

For $A, B \subset S$, define

$$R(x, A) = \mathsf{P}_x\{T_A < \infty\} \tag{24}$$

$$H(x, A) = \mathsf{P}_x\{S_A < \infty\} \tag{25}$$

$$O(x, A) = \mathsf{E}_x O_A \tag{26}$$

$$P^n_{!A}(x, B) = \mathsf{P}_x\{X_n \in B, T_A \geq n\}. \tag{27}$$

These are the return time probabilities, hitting probabilities, expected occupation times for the set A and taboo probabilities for the set B avoiding A.

Note that

$$R(x, A) = \sum_{n=1}^{\infty} P^n_{!A}(x, A). \tag{28}$$

Definition 24. All of the random variables in the last item are stopping times, meaning that $\{T = n\} \in \sigma(X_0, \ldots, X_n)$ for every n.

Let $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ be the history of the chain up to time n. For a stopping time T, we can define the information in the chain up to time T – the history up to time T – as a σ-field $\mathcal{F}_T$ defined as follows:

$$\mathcal{F}_T = \{A \in \mathcal{F} \text{ such that } A \cap \{T = n\} \in \mathcal{F}_n, \text{ for all } n \in \mathbb{Z}_{\oplus}\}. \tag{29}$$

Theorem 25. *The Strong Markov Property*

For a countable-state Markov chain and a bounded, measurable function h on sample paths,

$$\mathsf{E}\left(h(X_{n+1}, X_{n+2}, \ldots) \mid \mathcal{F}_T\right) 1\{T < \infty\} = \mathsf{E}_{X_T}(h(X_1, X_2, \ldots) 1\{T < \infty\}, \tag{30}$$

for all $n \geq 0$, where E_{X_T} corresponds to a chain whose initial distribution on $\mathcal{S}$ is the distribution of X_T. This is called the *Strong Markov Property.*