M464 - Probability 2 - Spring 2014 - Study Guide
Enrique Areyan.

CHAPTER 3: Markov Chains: Introduction.

Definition: A MARKOV process $\langle X_t ; t \geq 0 \rangle$ [not a set, so order matters]
satisfies: $\forall \underbrace{s_1, \ldots, s_j > t}_{\text{future}} ; \forall \underbrace{s_1', \ldots, s_k' < t}_{\text{past}} , \forall x_t ; x_{s_1}, \ldots, x_{s_j}, x_{s_j'}, \ldots, x_{s_k'}$:

$$P[X_{s_1} = x_{s_1}, \ldots, X_{s_j} = x_{s_j} \mid X_t = x_t, X_{s_1'} = x_{s_1'}, \ldots, X_{s_k'} = x_{s_k'}] =$$
$$P[X_{s_1} = x_{s_1}, \ldots, X_{s_j} = x_{s_j} \mid X_t = x_t].$$ [present is the only time that matters]

For a discrete-time chain (time index set is $\mathbb{N} = \{0, 1, 2, \ldots\}$), it is equivalent to require $\forall n \geq 0 : \forall x_1, \ldots, x_{n+1}$

$$P[X_{n+1} = x_{n+1} \mid X_n = x_n, X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] = P[X_{n+1} = x_{n+1} \mid X_n = x_n]$$

this is the MARKOV property of the MARKOV Chain.

Definition: we call the values of X_t states of the MARKOV Chain.
We usually take the state space to be $\mathbb{N} = \{0, 1, 2, \ldots\}$.
The conditional probabilities $P(X_{n+1} = j \mid X_n = i)$ are called the transition probabilities.
We usually consider only chains where these transition probabilities do not depend on n, called stationary transition probabilities.
We then write: $\boxed{P_{i,j} = P(X_{n+1} = j \mid X_n = i).}$
We call the matrix $[P_{ij}]_{i,j \in \mathbb{N}}$ the transition Probability Matrix.
This is a stochastic matrix: each row sums to 1 and has non-neg. entries.

---

n-step transition Probabilities: Write $P_{i,j}^{(n)} := P(X_{m+n} = j \mid X_m = i)$, (any $m \geq 0$)

THEOREM 2.1: The matrix $[P_{i,j}^{(n)}]$ is the n^{th} power of $[P_{i,j}]$ □.

If the initial prob. distribution is given as a row vector $[p_i]$, then the distribution at time n is given by: $[p_i][P_{ij}]^n$.
We can see this by the law of total prob:

$$P(X_n = j) = \sum_i Pr\{X_n = j, X_0 = i\} = \sum_i Pr\{X_n = j \mid X_0 = i\} Pr\{X_0 = i\} = \sum_i P_{ij}^{(n)} p_i$$

FIRST-STEP Analysis.

We use. If B is partitioned as $\{B_i\}$, Then: ← finite.

$$P(A|B) = \sum_i P(A|B_i, B)\, P(B_i|B) = \sum_i P(A|B_i)\cdot P(B_i|B) \quad \text{AND}$$

$$E[X|B] = \sum_i E[X|B_i, B]\, P(B_i|B) = \sum_i E[X|B_i]\, P(B_i|B).$$

Given a MARKOV CHAIN with probability matrix $\mathbf{P}$, where there is at least one absorbing state, define:

$$T = \min\{n \geq 0;\ X_n = \text{absorbing state } 1 \text{ OR } X_n = \text{absorbing state } 2\}$$

then X_T is the state of absorption. Now define:

$$u_i = \Pr\{X_T \overset{j \text{ absorbing state } (1 \text{ or } 2)}{=} j \mid X_0 = i\}.$$

Some u_i will be zero (that of the other absorbing state) some will be one (that of the absorbing state under consideration). By law of total prob.

$$u_1 = \Pr\{X_T = j \mid X_0 = 1\} = \sum_i \Pr\{X_T = j \mid X_0 = 1, X_1 = i\}\ \Pr\{X_1 = i \mid X_0 = 1\}$$

$$= \sum_i \Pr\{X_T = j \mid X_1 = i\}\ \Pr\{X_1 = i \mid X_0 = 1\} \quad \text{MARKOV Property.}$$

$$= u_1 P_{1,1} + u_2 P_{12} + u_3 P_{13} + \dots + u_n P_{1n}$$

Just replace the values from the appropiate row of $\mathbf{P}$.
You will get a system of equations that is solvable.

For expected time of absorption: Define

$$v_i = E[T \mid X_0 = i].$$ Some v_i will be zero (already absorb). (suppose state 1 is not absorbing)

Again, by law of total prob:

$$v_1 = E[T \mid X_0 = 1] = \sum_i E[T \mid X_0 = 1, X_1 = i]\ \Pr\{X_1 = i \mid X_0 = 1\}$$

$$= \sum_i E[T \mid X_0 = 1, X_1 = i]\ P_{1,i}$$

$$= 1 + v_1 P_{1,1} + v_2 P_{1,2} + v_3 P_{1,3} + \dots + v_n P_{1,n}$$

Add a one to account for the fact that you have to wait at least one more time step for absorption.

For a finite-state M.C., we call state i transient if $P_{i,i}^{(n)} \to 0$ as $n \to \infty$.

CHAPTER 4: Long-Run behavior of M.C.s.

Definition: A transition probability P on a finite number of states is called regular if $\exists k$ such that P^k has only positive entries: $\exists k: \forall_{i,j}: P_{ij}^{(k)} > 0$, i.e., there is a path of k transitions from i to j that all have positive probabilities.

THEOREM 1.1: Let P be regular. Then $\pi_j := \lim_{n \to \infty} P_{ij}^{(n)}$ exists for all i,j, does not depend on i, and is strictly positive.

If $\pi := (\pi_0\ \pi_1 \dots \pi_N)$, then π is the unique solution to $\pi P = \pi$, $\sum_{j=0}^{N} \pi_j = 1$.

To be regular, it is sufficient (but not necessary) that $\forall_{i,j}: \exists$ path from i to j of positive probabilities, AND $\exists i: P_{ii} > 0$.

Definition: Doubly Stochastic Matrix: all columns sum to one AND all rows sum to one. All entries ≥ 0.

If P is regular AND doubly stochastic then all long-run prob. are $\frac{1}{N}$, where N = number of states.

Long-run Mean fraction in states:

If P is regular, $\frac{1}{m} \sum_{k=0}^{m-1} P_{ij}^{(k)} \longrightarrow \pi_j$; that is, π_j is also the long-run mean fraction of time spent in state j.

Classification of states: Definitions:

- states i,j. Then j is accessible from i if $\exists n \geq 0$ s.t $P_{ij}^{(n)} > 0$
- j and i communicate ($i \leftrightarrow j$) if i and j are accesible from each other.
- this breaks the state space into equivalence classes (communicating classes)
- the Markov chain is irreducible if for all i,j: $i \leftrightarrow j$.

Periodicity: Given a state i, define the period of state i to be:

$$d(i) := \gcd\{n \geq 1 : P_{ii}^{(n)} > 0\}$$

theorem: (a) if $i \leftrightarrow j$ then $d(i) = d(j)$. (Period is a class property).

(b) For each state i: $\exists N(i)$ s.t. if $n \geq N$ then $P_{ii}^{(n d(i))} > 0$

(c) If $P_{ji}^{(m)} > 0$ then $\exists N$ s.t. if $n \geq N$ then $P_{ji}^{(m + n d(i))} > 0$

Definition: If all states in M.C. have period 1, we call the chain aperiodic.

Recurrence and transient states: Given state i, define the Probability of first returning to state i at the n^{th} step to be:

$$f_{ii}^{(n)} = P_r\{X_n = i, X_0 \neq i, X_1 \neq i, \ldots, X_{n-1} \neq i \mid X_0 = i\}$$ AND, $f_{ii}^{(0)} = 0$ for all i.

Note $f_{ii}^{(1)} = P_r\{X_1 = i \mid X_0 = i\} = P_{ii}$

$$f_{ii}^{(n)} = \sum_{k=0}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}, \quad n \geq 1.$$

Define: $f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} = P_r\{\underbrace{\text{return to } i \text{ eventually}}_{X_n = i \text{ for some } n \geq 1} \mid X_0 = i\}$

A state i is called recurrent if $f_{ii} = 1$, otherwise i is transient.

Let M = # of times that, starting from i, the process return to i. [# of returns could be zero]

$M \sim \text{Geometric}(1 - f_{ii}) - 1$

thus $E[M \mid X_0 = i] = \frac{1}{1 - f_{ii}} - 1 = \boxed{\frac{f_{ii}}{1 - f_{ii}}}$. this makes sense only if $f_{ii} < 1$. Otherwise the state is recurrent and will be visited infinitely many times (with probability 1). In this case $\sum_n P_{ii}^{(n)} = \infty$.

theorem: A state i is recurrent iff $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$.

Equivalently, i is Transient iff $\sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty$.

Corollary: If $i \leftrightarrow j$ and i is recurrent, then j is recurrent. (recurrence is a class property).

BASIC LIMIT THEOREM OF MC'S: If state i is recurrent, we may define the R.V. $R_i := \min\{n \geq 1 ; X_n = i\}$. Now, when $X_0 = i$, this R.V is the first return to i. Its distribution is $f_{ii}^{(n)}$ $(n \geq 1)$.

Hence, $E[R_i \mid X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^{(n)} := m_i$ [mean time to return to state i]

<u>theorem</u>: For a recurrent, irreducible aperiodic M.C. and all i,j, we have:

$\lim_{n\to\infty} P[X_n=j] = \lim_{n\to\infty} P_{ji}^{(n)} = \frac{1}{m_i} =: \pi_j$ [these do not need to add to one]

the recurrent states that are not positive recurrent are called <u>null recurrent</u> [Positive recurrent means $\overline{\lim_{n\to\infty}} P_{ii}^{(n)} > 0$] this is a class property.

If a M.C. is aperiodic, irreducible and recurrent then then it is positive recurrent iff all $m_i < \infty$

<u>THEOREM</u>: For a positive recurrent, irreducible, aperiodic M.C., there is a unique <u>stationary probability distribution</u> $\underline{\pi}$:

$$\underset{\sim}{\pi} > 0, \quad \underset{\sim}{\pi} \cdot \mathbb{1}^T = 1, \quad \underline{\pi} \cdot P = \underline{\pi} \leftarrow \text{stationary}.$$

We have for all i; $\pi_i = \frac{1}{m_i}$.

Conversely, an irreducible aperiodic M.C. with a stationary probability distribution is positive recurrent.

Suppose that an irreducible recurrent M.C. $X_0, X_1, X_2, \ldots$, has period $d \geq 1$ then $\lim_{n\to\infty} P_{ii}^{(n)}$ either $=0$ or (if $d>1$) does not exist. thus, let $Y_n := X_{nd}$ then $Y_0, Y_1, Y_2, \ldots$, is a M.C. with transition matrix $\underset{\sim}{P}^d$, where $\underset{\sim}{P}$ is the transition matrix on X, and so $\underset{\sim}{P}^d$ is aperiodic, recurrent, but have d com. classes still, $\lim_{n\to\infty} P_{ii}^{(nd)} = \lim_{n\to\infty} P[Y_n = i \mid Y_0 = i]$ exists by restricting to the states that communicate with i for $\langle Y_n \rangle$. the mean time for $\langle Y_n \rangle$ to return to i is d times the mean time m_i MEAN TIME for $\langle X_n \rangle$ to return to i is d times the mean time for $\langle Y_n \rangle$: $\lim_{n\to\infty} P_{ii}^{(nd)} = \frac{d}{m_i}$. The long-run fraction of time is $\frac{1}{m_i}$ (SAME AS APERIODIC CASE). If $\langle X_n \rangle$ pos. recurrent then $\pi_i := \frac{1}{m_i}$ unique stationary prob. distribution.