Prof. H. Bensmail

Stat-664: Adv Stat Infer

Spring Semester 2006

Chapter 6:

An overview of Discrete Markov Chain

March 1, 2006

Introduction

Definition:
A finite state, discrete Markov Chain is a sequence of random variables X_n $(n = 0, 1, \ldots\ldots)$ where X_n take values in the finite set $S = (0, 1, \ldots, N-1)$ and are called the states of the Markov chain.

n : is the stages or time periods.

if $X_n = i$: process is said to be in state i at time n.

A Markov must satisfy the following property:

$$P(X_n = j | X_{n-1} = i, X_{n-2} = k, \ldots, X_0 = m) =$$
$$P(X_n = j | X_{n-1} = i) = p^{(1)}(i,j), \quad i,j,k,\ldots,m \in S$$

It means that:
The conditional distribution at time n, given all the past states $X_0 = m, \ldots, X_{n-1} = i$, only depends on the immediately preceding state, $X_{n-1} = i$.

Particular case: a sequence of independent random variables.

$p_{i,j}$ is the transition probability
It is the probability that the chain at time n is in the state j, given that at time $n-1$ it was in state i.

State of the system:
A matrix P

$$P = \begin{bmatrix} p(0,0) & \cdots & p(0,N-1) \\ \vdots & & \vdots \\ p(N-1,0) & \cdots & p(N-1,N-1) \end{bmatrix}$$

- describes the $(N \times N)$ probability of transition of the chain in *one* time period.
- The probability of a transition from a given state to another depends on the two states and not on time.
- The $(i+1)th$ row of P is the probability distribution of values of X_n under the condition that $X_{n-1} = i$.
- Every entry of P satisfies $p(i,j) \geq 0$, and every row of P satisfies:

$$\sum_j p(i,j) = 1$$

Example:
When $S = \{0,1,2\}$

$$P = \begin{bmatrix} p(0,0) & p(0,1) & p(0,2) \\ p(1,0) & p(1,1) & p(1,2) \\ p(2,0) & p(2,1) & p(2,2) \end{bmatrix}$$

Question:
Calculate the probability that a random variable will be in state 2, at time $2+m,$ given that it was in stage 1 at time m:

$$p^{(2)}(1,2) = P(X_{2+m} = 2|X_m = 1) \quad (1)$$

$$\sum_{j=0}^{2} P(X_{2+m} = 2, X_{1+m} = j|X_m = 1) = \quad (2)$$

$$\sum_{j=0}^{2} P(X_{1+m} = j|X_m = 1)P(X_{2+m} = 2|X_{1+m} = j) \quad (3)$$

$$p(1,0)p(0,2) + p(1,1)p(1,2) + p(1,2)p(2,2) \quad (4)$$

Last line can be recognized as the product of row 2 of P (associated with state 1) and column 3 of P (associated with state 2).

In general:

$$p^{(n)}(i,j) = P(X_n = j|X_0 = i) =$$
$$P(X_{n+k} = j|X_k = i)$$

The element $p^{(n)}(i,j)$ is the $(i+1,j+1)$ entry of P^n (Chapman-Kolmogorov).

$$\begin{aligned} p^{(m+n)}(i,j) &= P(X_{m+n} = j|X_0 = i) \\ &= \sum_{k=0}^{N-1} P(X_{m+n} = j, X_m = k|X_0 = i) \\ &= \sum_{k=0}^{N-1} P(X_{m+n} = j|X_m = k)P(X_m = k|X_0 = i) \\ &= \sum_{k=0}^{N-1} p^{(m)}(i,k)p^{(n)}(k,j) \end{aligned}$$

The matrix summarizing the equation above by:

$$P^{m+n} = P^m P^n$$

Long-Term behavior:

Let $\pi^{'(n)}$ be the N-dimensional row vector denoting the *probability distribution* of X_n.
The $i^{(th)}$ component of $\pi^{(n)}$ is

$$\pi^{(n)}(i) = P(X_n = i),\ i \in S$$

When $n = 0$,
$\pi^{(0)}$: the initial probability distribution of the chain.

$$\begin{aligned} P(X_n = j) &= \sum_{k=0}^{N-1} P(X_n = j|X_{n-1} = k)P(X_{n-1} = k) \\ &= \sum_{k=0}^{N-1} p(k,j)P(X_{n-1} = k), j = 0,1,\ldots,N-1 \qquad (*) \end{aligned}$$

$P(X_n = j)$ is the $(j+1)th$ element of $\pi^{(n)}$
$(*)$ is the product of the row vector $\pi^{'(n-1)}$ with column $(j+1)$ of the transition matrix P.

So we have:

$$\pi^{'(n)} = \pi^{'(n-1)}P$$

and

$$\pi^{'(1)} = \pi^{'(0)}P, \pi^{'(2)} = \pi^{'(1)}P = \pi^{'(0)}PP = \pi^{'(0)}P^2$$

and so on....then we have:

$$\pi^{'(n)} = \pi^{'(0)}P^n$$

Conclusion*:*

- ***The random evolution of the Markov chain is completely specified in terms of the distribution of the initial state and the transition probability matrix*** P.
- ***So given the initial probability distribution and the transition probability matrix, it is possible to describe the behavior of the process at any specified time period*** n.

Stationary Distribution:

We have the following result:
The *convergence* ***of the chain to a limiting distribution independent of any legal starting distribution***.

If P^n converges to some invariant matrix π', then this should verify:

$$\begin{aligned} \pi' &= \lim_{n\to\infty} \pi^{'(0)}P^n \\ &= (\lim_{n\to\infty} \pi^{'(0)}P^n)P \\ &= \pi'P \end{aligned}$$

The distribution π is said to be a ***stationary distribution*** (also known as the invariant or ***equilibrium distribution***) if it satisfies

$$\pi' = \pi' P$$

$$or$$

$$\pi(j) = \sum_{i=0}^{N-1} \pi(i)p(i,j)$$

<u>Example:</u>

Three-state space Markov chain:
Suppose $S = \{0,1,2\}$ with the transition matrix

$$P = \begin{bmatrix} p(0,0) & p(0,1) & p(0,2) \\ p(1,0) & p(1,1) & p(1,2) \\ p(2,0) & p(2,1) & p(2,2) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Meaning:
The probability distribution of the values of X_{n+1} given $X_n = 1$ are:

$$p(1,0) = \tfrac{1}{4}, \quad p(1,1) = \tfrac{1}{2}, \quad p(1,2) = \tfrac{1}{4}$$

The element $p(1,2)$ is the probability that the system moves from states 1 to state 2 in *one* transition.

we have

$$P^4 = \begin{bmatrix} 0.2760 & 0.4653 & 0.2587 \\ 0.2326 & 0.4485 & 0.3189 \\ 0.1725 & 0.4251 & 0.4024 \end{bmatrix}$$

and $p^4(1,2) = 0.3189$ is the probability of moving from state 1 to state 2 in *four* transitions.
If the starting probability distribution of the chain is

$$\pi'^{(0)} = (0\ 1\ 0)$$

then we have:

$$\pi'^{(1)} = (0.2500\ \ 0.5000\ \ 0.2500)$$
$$\pi'^{(2)} = (0.2500\ \ 0.4583\ \ 0.2917)$$
$$\pi'^{(4)} = (0.2396\ \ 0.4514\ \ 0.3090)$$
$$\pi'^{(5)} = (0.2326\ \ 0.4485\ \ 0.3189)$$

So the system converges to the stationary distribution

$$\pi' = (0.2222 \;\; 0.4444 \;\; 0.3333)$$

We can obtain this uniquely using

$$\pi' = \pi' P \text{ and } \pi(j) = \sum_{i=0}^{N-1} \pi(i) p(i,j)$$

Suppose that

$$\pi' = (\pi(0) \;\; \pi(1) \;\; \pi(2))$$

we have

$$\pi(2) = 1 - \pi(1) - \pi(0)$$

The system of equation to be solved is:

$$\frac{\pi(0)}{2} + \frac{\pi(1)}{4} = \pi(0)$$

$$\frac{\pi(0)}{2} + \frac{\pi(1)}{2} + \frac{1 - \pi(1) - \pi(0)}{3} = \pi(1)$$

the unique solution is

$$\pi(0) = 2/9, \; \pi(1) = 4/9, \; \pi(2) = 1 - \pi(1) - \pi(0) = 3/9$$

Aperiodicity and Irreducibility

Aperiodicity

To have convergence and uniqueness, the finite chain is required to be aperiodic and irreducible.

Consider the chain which has this only possible transitions are:

$1 \to 2, \quad 2 \to 3 \quad 3 \to 1$

Which means that the states are reached every 3 movements of the chain.

$2 \to 3 \quad 3 \to 1 \quad 1 \to 2$ (time 3) $\quad 2 \to 3 \quad 3 \to 1 \quad 2 \to 3$ (time 6)

$3 \to 1 \quad 1 \to 2 \quad 2 \to 3$ (time 9)

So we say that this state has period equal to 3.

Definition:

If the period of state j of the chain is d, it means that $p^{(n)}(j,j) = 0$ whenever n is not divisible by d, and the largest integer with this property.

or

The period d of a chain is the greatest common divisor of all n such that

$p^{(n)}(j,j) > 0$.

An aperiodic state has period 1, or alternatively, a state j is aperiodic if $p^{(n)}(j,j) > 0$ for all sufficiently large n.

So a sufficient condition for a state to have period 1 is

$$P(X_n = j|X_0 = j) \text{ and } P(X_{n+1} = j|X_0 = j) > 0$$

$$p^{(n)}(j,j) > 0 \text{ and } p^{(n+1)}(j,j) > 0$$

So when $p(j,j) = P(X_n = j|X_{n-1} = j) > 0$ for all j

Irreducibility:

***Definition**:*

A chain is called irreducible, if each state is reachable from every other state in a finite number of transitions

Irreducibility means that for every pair of states (i,j),

$$p^{(k)}(i,j) = P(X_{n+k} = j|X_n = i) > 0 \text{ for } k \geq 0$$

***Definition 1**:*

A chain that is irreducible with period d has a transition probability matrix with d eigenvalues with absolute value 1.

***Definition 2**:*

A chain which is aperiodic and irreducible is said to be: ergodic

***Definition 3**:*

An ergodic Markov chain converges to a unique stationary distribution.

Example 1:

Consider a Markov chain with the following transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Calculate: $P^{(1)}, P^{(2)}, P^{(3)}, P^{(4)}, \ldots$.

if n is <u>even</u>:

$$P^n = \begin{bmatrix} 0.25 & 0 & 0.50 & 0 & 0.25 \\ 0 & 0.50 & 0 & 0.50 & 0 \\ 0.25 & 0 & 0.50 & 0 & 0.25 \\ 0 & 0.50 & 0 & 0.50 & 0 \\ 0.25 & 0 & 0.50 & 0 & 0.25 \end{bmatrix}$$

and is n is <u>odd</u>:

$$P^n = \begin{bmatrix} 0 & 0.50 & 0 & 0.50 & 0 \\ 0.25 & 0 & 0.50 & 0 & 0.25 \\ 0 & 0.50 & 0 & 0.50 & 0 \\ 0.25 & 0 & 0.50 & 0 & 0.25 \\ 0 & 0.50 & 0 & 0.50 & 0 \end{bmatrix}$$

The Markov chain is irreducible because $p^{(n)}(i,j) > 0$ for all i,j
Periodic with period $d = 2$
It means that starting in state 1, for example, $p^n(1,1) > 0$ at times $n = 2,4,6,8,\ldots.$ and the <u>greatest common divisor</u> of the values that n can take is 2.

The unique stationary distribution is

$$\begin{aligned} \pi &= \lim_{n\to\infty} \frac{1}{2}(\pi^{(n)} + \pi^{(n+1)}) \\ &= [0.125 \;\; 0.25 \;\; 0.25 \;\; 0.25 \;\; 0.125] \end{aligned}$$

we can verify that the eigenvalues of P are $-1, 0, 1, -1/\sqrt{2}, 1/\sqrt{2}$
and there are $d = 2$ eigenvalues with absolute value 1

<u>Example 2</u>:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

this system has 5 stationary distributions so for large n :

Calculate $P^{(1)}, P^{(2)}, \ldots P^{(n)}$

$$P^n = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.75 & 0 & 0 & 0 & 0.25 \\ 0.50 & 0 & 0 & 0 & 0.50 \\ 0.25 & 0 & 0 & 0 & 0.75 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The rows of P^n represent the five stationary distributions, and each of these satisfy $\pi' = \pi' P$

Example 3:

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{3}{24} & \frac{16}{24} & \frac{5}{24} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{5}{6} \end{bmatrix}$$

$$P^n = \begin{bmatrix} 0.25 & 0.75 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 & 0 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \end{bmatrix}$$

The chain split into two sub-chain, each converges to an equilibrium distribution and one can not move from state spaces $\{0,1\}$ to state $\{2,3,4\}$.

Reversible Markov chain:

Consider a Markov chain with state space S that converges to an invariant distribution π.
Let $x \in S$ denote the current state of the system.
Let $y \in S$ denote the current state at the next step.
Let $p(x,y)$ be the probability of a transition from x to y
Let $p(y,x)$ be the probability of a transition from y to x.

A Markov chain is said to be reversible if it satisfies the condition:

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

then we have:

$$\pi(x) = \sum_{y \in S} \pi(y)p(y,x)$$

$$\begin{aligned} \pi(x)p(x,y) &= P(X_n = x)P(X_{n+1} = y|X_n = x) \\ &= P(X_n = x, X_{n+1} = y), \text{ for all } x, y \in S. \end{aligned}$$

So the reversibility condition can also be written as

$$P(X_n = x, X_{n+1} = y) = P(X_n = y, , X_{n+1} = x) \text{ for all } x, y \in S.$$

The reversibility will be important in deriving the appropriate transition kernels function for the Metropolis algorithm or other Markov chain.