HGEN 48600/STAT 35450: Lecture 5

Lecturer: John Novembre Scribe: Chi-Chun Liu

Note: *These lecture notes are still rough, and have only have been mildly proofread.*

Discrete Markov Chain

$$X_1 \longrightarrow X_2 \longrightarrow X_3 \longrightarrow X_n$$

Time homogeneous markov chain:

$P(X_{n+1} = j | X_n = i) = P(X_2 = j | X_1 = i) = P_{ij}$, for all n.
Time homogeneous markov chains have properties that make them easier to analyze.

Discrete Markov Chain Examples:

Random walk :

Given a current state $X_i = k$, X_{i+1} is either $k + 1$ or $k - 1$ with equal probability.
In terms of transition matrix,
$P_{i,i-1} = \frac{1}{2}$. $P_{i,i+1} = \frac{1}{2}$. $P_{i,i} = 0$, and otherwise 0.
(More in Ross example 4.18)

Jukes Cantor Model

Given a current state of a single site, it has probability μ of substitution, with the same probability of mutating into one in other three states.

$$M = \begin{array}{c c} & \begin{array}{cccc} A & C & T & G \end{array} \\ \begin{array}{c} A \\ C \\ T \\ G \end{array} & \begin{pmatrix} 1-\mu & \mu/3 & \mu/3 & \mu/3 \\ \mu/3 & 1-\mu & \mu/3 & \mu/3 \\ \mu/3 & \mu/3 & 1-\mu & \mu/3 \\ \mu/3 & \mu/3 & \mu/3 & 1-\mu \end{pmatrix} \end{array}$$

Wright-Fisher Model

We have N diploid individuals, and there's no overlap between generations.
X_i: number of copies of allele A.
$X_i \sim Binom(2N, \frac{X_{i-1}}{2N})$

$$P = \begin{bmatrix} \quad \end{bmatrix} \in R^{2N \times 2N}$$

Each row of P is a binomial distribution. $P_{ij} = \binom{2N}{j}(\frac{i}{2N})^j(1 - \frac{i}{2N})^{2N-j}$

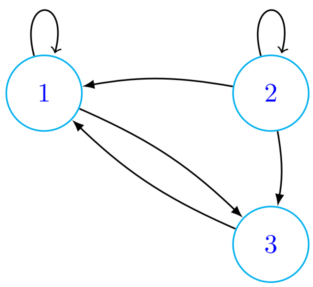
Useful formula

Chapman-Kolmogorov:
Let $P_{ij}^{(n)} = P(X_{n+k} = j, X_k = i)$.
$P^{(n+m)} = p^{(n)}p^{(m)}$, e.g. $P^{(2)} = p^{(1)}p^{(1)}$.

$P(X_1 = j) = \sum_i P(X_0 = i)P(X_1 = j|X_0 = i)$, where $P(X_0 = i)$ is initiated from π_0.
e.g. $\pi_1^T = \pi_0^T P$, $\pi_{10}^T = \pi_0^T P^{10}$.

Forward Kolmogorov:
$\pi^n = \pi^{(n-1)}P$

Classification of states



$$P = \begin{bmatrix} .8 & 0 & .2 \\ .2 & .2 & .6 \\ .2 & 0 & .8 \end{bmatrix}$$

f_i = the probability of reentering at state i given we start at state i.
$f_i = P(T_i < \infty | X_0 = i)$, where $T_i = \{min\ n > 0 : X_n = i\}$.
$f_i = 1 \rightarrow$ state i is recurrent.
$f_i < 1 \rightarrow$ state i is transient.

Remark: with the markov property, when recurrence happens, it re-initialize the chain. Hence the process will reenter the recurrent states infinite times.

Also let $N_j = \min\{n > 0 : X_n = j\}$, and $m_j = E[N_j|X_0 = j]$. The chain is **positively recurrent** if $m_j < \infty$, and otherwise **null recurrent**.

Ergodic Chain

a. irreducible: single class of states (Within a class, states communicate with each other)
b. all states are aperiodic (periodic: chain can return to the state only periodically. It is also a class property)
c. all states are recurrent

An ergodic chain with finite state space has a single stationary distribution.

Global Balance Equation

$\pi^T = \pi^T P$
Also recall the R vignette. $P^{(n)}$ with $n \rightarrow \infty$ would have a stationary distribution in every row.

Aside: matrix power

Suppose P is diagonalizable, we perform EVD and let $P = V\Lambda V^{-1}$, where the columns of V are right eigenvectors of P, and rows of V^{-1} are left eigenvectors of P. Then $P^n = V\Lambda^n V^{-1}$.

EVD is useful for finding stationary states for markov chain. π, the stationary state could be found normalizing the left eigenvector associated with eigen value 1. (Or it can be obtained from $\pi P = \pi$ s.t. $\sum_i \pi = 1$)

Note that in **R** the eigen command computes right eigenvectors, one can inverse V to obtain left eigenvectors.

Time Reversible Markov Chain

$Q_{ij} = P(X_m = j|X_{m+1} = i) = \frac{P(X_m=j, X_{m+1}=i)}{P(X_{m+1}=i)} = \frac{P(X_{m+1}=i|X_m=j)P(X_m=j)}{P(X_{m+1}=i)} = \frac{\pi_j P_{ji}}{\pi_i}$.
When $Q_{ij} = P_{ij}$, or $\pi_i P_{ij} = \pi_j P_{ji}$ we say its time reversible.

In plain text, it says for all states i and j, the rate at which the process goes form i to j is equal to the rate at which it goes from j to i.

Reference

1. Discrete-Time Markov Chains, five-minute stats

2. Ch4, Ross