LECTURE NOTES: Discrete time Markov Chains (2/24/03; SG)

A discrete time Markov Chains is a stochastic dynamical system in which the probability of arriving in a particular state at a particular time moment depends only on the state at the previous moment. That is,

- states are discrete: $i = 0, 1, 2, \ldots$
- time is discrete: $t = 1, 2, \ldots$
- probabilities p_{ij} of transition from state i to state j in one time step are constant, i.e., they do not depend on time and do not depend on how the system got in state i ("Markov property").

Island example

- Consider an island that may be repeatedly colonized by a certain species which also may go extinct.
- State variable:

$$X(t) = \begin{cases} 0 & \text{if island is empty at time } t, \\ 1 & \text{if island is occupied at time } t. \end{cases}$$

- Dynamic rule:

$$\text{If } X(t) = 0, \text{then } \begin{cases} X(t+1) = 0 & \text{with prob. } 0.2, \\ X(t+1) = 1 & \text{with prob. } 0.8. \end{cases}$$

$$\text{If } X(t) = 1, \text{then } \begin{cases} X(t+1) = 0 & \text{with prob. } 0.1, \\ X(t+1) = 1 & \text{with prob. } 0.9 \end{cases}$$

That is, the probability of colonization is $p_{col} = 0.2$ and the probability of extinction is $p_{ext} = 0.1$.

QUESTION: what is the long-term behavior of the system?

Matlab code

- set parameters: number of generations T, p_{col} and p_{ext}
- set initial conditions: $X(1) = 0$ (island is empty)

- set update rule
- set graphics: $plot(X)$
- print frequency of being occupied:
$$X_{mean} = mean(X)$$
- play with the model and look for patterns
- QUESTIONS: can we predict $X(t)$? What can be predicted?

Analytical theory

Define the probabilities

$$\begin{aligned} x_0(t) &= Pr(X(t) = 0), \\ x_1(t) &= Pr(X(t) = 1). \end{aligned}$$

Recall the Law of Total Probability: Let B_i be a set of mutually exclusive (no pair intersect) and collectively exhaustive (the union is the full set) events. Then

$$Pr(A) = \sum_i Pr(A|B_i)Pr(B_i)$$

where $Pr(B_i)$ is the probability of B_i and $Pr(A|B_i)$ is the probability of A conditioned on B_i.

Using the Law of Total Probability

$$\begin{aligned} Pr(X(t+1) = 0) =& Pr(X(t+1) = 0|X(t) = 0)\ Pr(X(t) = 0) + \\ & Pr(X(t+1) = 0|X(t) = 1)\ Pr(X(t) = 1), \end{aligned}$$

$$\begin{aligned} Pr(X(t+1) = 1) = & Pr(X(t+1) = 1|X(t) = 0)\ Pr(X(t) = 0) + \\ & Pr(X(t+1) = 1|X(t) = 1)\ Pr(X(t) = 1), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} x_0(t+1) &= p_{00}\ x_0(t) + p_{01}\ x_1(t), \\ x_1(t+1) &= p_{10}\ x_0(t) + p_{11}\ x_1(t), \end{aligned}$$

which, in turn, can be rewritten in matrix notation as

$$(x_0(t+1), x_1(t+1)) = (x_0(t), x_1(t)) \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix}$$

Finally, let the distribution vector be

$$x(t) = (x_0(t), x_1(t))$$

and the transition matrix be

$$P = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix}.$$

Then the dynamics of $x(t)$ are described by

$$x(t+1) = x(t)P.$$

Note that

$$\begin{aligned} x(1) &= x(0)P, \\ x(2) &= x(1)P = x(0)P^2, \\ x(3) &= x(2)P = x(0)P^3, \\ &\dots \end{aligned}$$

Thus

$$x(t) = x(0)P^t,$$

where vector $x(t)$ characterizes the initial probabilities. For example, if the island is initially empty, then

$$x(0) = (1, 0).$$

Matlab code

- set parameters: number of generations T, probabilities p_{col}, p_{ext} and the transition matrix P
- set initial distribution: $x = [1\ 0]$ (island is empty)
- set update rule: $x(t,:) = x(1,:) * P^t$
- set graphics:

$$plot(1:T, x(:,1), 1:T, x) :, 2)$$

- print asymptotic frequency of being occupied: $x(T, 2)$
- play with the model and look for patterns
- compare results with stochastic simulations

Random genetic drift

Consider a randomly mating diploid population with size N. Assume there is a single locus with alleles **A** and **a**. The population state is characterized by the number of alleles **A**:

$$X = 0, 1, 2, \ldots 2N.$$

Let

$$x_i(t) = Pr(X = i)$$

be the probability that the population has i alleles **A** at generation t. Can we predict the dynamics of $x_i(t)$ using a Markov chain model?

Let

$$q_i = \frac{i}{2N}$$

be the current frequency of **a**, and

$$1 - q_i$$

be the current frequency of **A**. Under the Fisher-Wright binomial scheme for random genetic drift the transition probabilities are

$$P_{ij} = \binom{2N}{j} q_i^j \; (1 - q_i)^{2N-j}$$

where $\binom{2N}{j}$ is the binomial coefficient.

Matlab code

- set parameters: number of generations T, population size N, the initial frequency $p0$
- set the transition matrix (the binomial coefficient is given by command

$$nchoosek(2 * N, j)$$

- set initial population
- set graphics: place

$$bar(1 : 2 * B, x(t, :))$$

  inside the loop $for\ t = 2 : T \ldots end$
- play with the model and look for patterns
- compare results with stochastic simulations

Asymptotic state

Dynamics equation

$$x(t+1) = x(t)P$$

has an equilibrium x that satisfies to matrix equation

$$x = xP.$$

The latter can be rewritten as

$$xP = \lambda x$$

with $\lambda = 1$. This shows that the equilibrium distribution is given by the **eigenvector** of matrix P corresponding to the **eigenvalue** $\lambda = 1$.

To find the eigenvectors and eigenvalues use command

$$[V, D] = eig(P')$$

(notice the transpose sign $'$). Type

$$help \quad eig$$

to get help. You will need to normalize the eigenvector, say k-th eigenvector, by doing

$$V(:, k)/sum(V(:, k))$$

Find the stationary distributions for the Island model and for the Random Genetic Drift model.

Absorption states: absorption probabilities

State i is an absorption state if $p_{ii} = 1$ (can't leave it once you get there).

In the Random Genetic Drift model there are two absorption states: $i = 0$ and $i = 2 * N$.

Assume there are two absorption states: O and Ω. Let π_i be the probability of absorption at state Ω rather than at state 0 if starting at state i. Then p_i can be found from a system of linear algebraic equations

$$\pi_0 = 0,$$
$$\pi_\Omega = 1,$$
$$\pi_i = \sum_j P_{ij}\pi_j$$

Using matrix notation, let

$$\pi = (\pi_1, \pi_2, \ldots, \pi_{\Omega-1}),$$

Let $\tilde{P}$ be a truncated matrix P (with rows and columns corresponding to the absorption states removed).

Let p_Ω be the column of P corresponding to Ω with the element corresponding to state 0 removed.

Then

$$\pi = \left(I - \tilde{P}\right)^{-1} p_\Omega,$$

where I is the identity matrix.

Absorption states: average absorption times

Let T_i be the average time to reach an absorption state if starting at state i. Then T_i can be found from a system of linear algebraic equations

$$T_0 = 0,$$
$$T_\Omega = 0,$$
$$T_i = 1 + \sum_j P_{ij} T_j$$

Using matrix notation, let

$$T = (T_1, T_2, \ldots, T_{\Omega-1}),$$

Then

$$T = \left(I - \tilde{P}\right)^{-1} J,$$

where $J = (1, 1, \ldots, 1$ is a vector of 1's.

Use Matlab code *fixation.m* to study fixation probabilities and fixation times in the Random Genetic Drift model.

Assignment

Assume there are S sites arranged along a line. Assume that an individual starts at site i and makes a step right or left with equal probability $m (< 1/2)$. Study this "random walk" using stochastic simulations and a Markov chain model. How long does it take to reach "absorbing" states 1 and S. Use several different values of S and m.