BioInformatics Summer Institute - Probability/Statistics Workshop

Instructor: **Dr. Silvia Heubach**

Workshop II – Markov Chains

Assignments:

4. Markov Chains

4.1 An $n \times n$ matrix with non-negative entries is **stochastic** if the entries in each row sum to 1. The matrix is **doubly stochastic** if each row and each column sums to 1. A Markov chain is **irreducible** if every state can be reached from every other state, i.e., its transition matrix is **regular**.

a) For each of the following matrices, determine whether it is stochastic, doubly stochastic, and/or regular.

i)
$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

ii)
$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$

iii)
$$P = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.75 & 0 \end{pmatrix}$$

iv)
$$P = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0.9 & 0 & 0.1 \\ 0.25 & 0.75 & 0 \end{pmatrix}$$

b) A state is called **absorbing** if the probability of leaving the state is 0. How can you identify absorbing states from the transition matrix?

4.2 Electrical usage during a summer month can be classified as normal, high or low. Past records have shown that the usage pattern changes according to the following matrix:

$$\begin{array}{c} \\ N \\ H \\ L \end{array} \begin{array}{c} \begin{array}{ccc} N & H & L \end{array} \\ \begin{pmatrix} 3/4 & 1/6 & 1/12 \\ 2/5 & 1/3 & 4/15 \\ 1/2 & 2/5 & 1/10 \end{pmatrix} \end{array}$$

Find the stationary vector for this matrix and interpret its meaning.

4.3 Voters often change their party affiliation in subsequent elections. In a certain district, Republicans stay Republicans with probability 0.8, while Democrats stay with their party with probability 0.9. Visualize the chain, derive the transition matrix and find the stationary distribution.

4.4 Suppose you have a DNA sequence with heterogeneous base composition, consisting of two states. State 1 generates an AT-rich sequence, whereas state 2 generates a CG-rich sequence. The DNA sequence stays in the same state with probability 0.99, and changes state with probability 0.01.

a) Visualize this process in a diagram.
b) Derive the transition matrix.
c) What is the probability of seeing the sequence 1-1-1-1-1-2-2-2-2-1-1?
d) If the sequence is equally likely to be in either state initially, what is the probability that $X_7 = 1$?
e) What happens to the chain in the long run?

4.5 Suppose you have one die. Define the *state* of the die to be the number showing on top. Instead of rolling the die, pass from one state to the next by tipping the die by 90^{o} in any direction with equal probability. Recall that numbers on opposite sides of a die add up to seven, so if the current state is 6, then the next state can be 2, 3, 4, or 5 (1 is at the bottom, and 6 cannot occur because the state has to change). This process defines a first order Markov chain.

a) Visualize this process in a diagram.
b) Derive the transition matrix.
c) Suppose the initial state is a 5. What is the probability that another five occurs on the fourth move? (Use the instructions for MatLab to perform the matrix operations).
d) Find the stationary distribution for this chain.
e) Experimentally find out after how many steps k the state has reached the stationary distribution, i.e., $\vec{p}^{(0)}P^k = \pi$ for
   i) $\vec{p}^{(0)} = (1\ 0\ 0\ 0\ 0\ 0)$
   ii) $\vec{p}^{(0)} = (\frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6})$
   iii) $\vec{p}^{(0)} = (\frac{1}{3}\ \frac{1}{3}\ 0\ \ 0\ \frac{1}{3}\ 0)$
   iv) $\vec{p}^{(0)} = (\frac{1}{2}\ \frac{1}{4}\ \frac{1}{8}\ \frac{1}{16}\ \frac{1}{32}\ \frac{1}{32})$.