Some Applications of Markov Chains

1. Squirrels

The American gray squirrel (*Sciurus carolinensis* Gmelin) was introduced in Great Britain by a series of releases from various sites starting in the late nineteenth century. In 1876, the first gray squirrels were imported from North America, and have subsequently spread throughout England and Wales, as well as parts of Scotland and Ireland.

Simultaneously, the native red squirrel (*Sciurus vulgaris* L.), considered the endemic subspecies, has disappeared from most of the areas colonized by gray squirrels. Originally, the red squirrel was distributed throughout Europe and eastward to northern China, Korea, and parts of the Japanese archipelago. During the last century, the red squirrel has consistently declined, becoming extinct in many areas of England and Wales, so that it is now confined almost solely to Northern England and Scotland. A few isolated red squirrel populations exist on offshore islands in southern England and mountainous Wales.

The introduction of the American gray squirrel continued until the early 1920s, by which time the gray squirrels had rapidly spread throughout England. By 1930 it was apparent that the gray squirrel was a pest in deciduous forests, and control measures were attempted. Once the pest status of the gray squirrel was recognized, national distribution surveys were undertaken. The resulting distribution maps clearly showed the tendency for the red squirrel to be lost from areas that had been colonized by the gray squirrel during the preceding 15 to 20 years.

Since 1973, an annual questionnaire has been circulated to foresters by the British Forestry Commission. The questionnaire concerns the presence or absence of the two squirrel species. It also includes questions on the changes of squirrel abundance, details of tree damage, squirrel control measures, and the number of squirrels killed. Using the data collected by the Forestry Commission, we wish to construct a model to predict the trends in
the distribution of both species of squirrels in Great Britain.

Several researchers have studied the British squirrel populations, notably Reynolds [3] and Usher et al. [5]. The annual Forestry Commission data has been summarized in the form of distribution maps reflecting change over a two-year period.

Usher et al., [5] used an overlay technique to extract data from the distribution maps. Each 10-km square on the overlay map that contained Forestry Commission land is classified into one of four states:

- **R:** only red squirrels recorded in that year.
- **G:** only gray squirrels recorded in that year.
- **B:** both species of squirrels recorded in that year.
- **O:** neither species of squirrels recorded in that year.

In order to satisfy the Markov assumption, squares that were present only in two consecutive years were counted. Counting the pairs of years, squares are allocated to any one of 16 classes, e.g., \( R \rightarrow R, R \rightarrow G, G \rightarrow G, B \rightarrow O, \) etc.

A summary of these transition counts for each pair of years from 1973–74 to 1987–88 is given in table 1 and is reprinted by permission of Blackwell Science Inc.

A frequency interpretation is required to employ the Markov chain analysis. If the

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>G</th>
<th>B</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2,529</td>
<td>35</td>
<td>257</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>61</td>
<td>733</td>
<td>20</td>
<td>91</td>
</tr>
<tr>
<td>B</td>
<td>282</td>
<td>25</td>
<td>4,311</td>
<td>335</td>
</tr>
<tr>
<td>O</td>
<td>3</td>
<td>123</td>
<td>310</td>
<td>5,930</td>
</tr>
</tbody>
</table>

Table 1: Red and Gray Squirrel Distribution Map Data for Great Britain.
entries in each column are totaled, the corresponding matrix entry is found by division. For example, column $R$ has a total $2,529 + 61 + 282 + 3 = 2,875$, so that the entry in the $R, R$ position is $2,529/2,875 \approx 0.8797$. Care must be taken when calculating these frequencies. Inappropriate rounding will violate the requirement that the columns sum to one. The transition matrix (rows and columns are in $R$, $G$, $B$, $O$ order) is

$$
T = \begin{pmatrix}
0.8797 & 0.0382 & 0.0527 & 0.0008 \\
0.0212 & 0.8002 & 0.0041 & 0.0143 \\
0.0981 & 0.0273 & 0.8802 & 0.0527 \\
0.0010 & 0.1343 & 0.0630 & 0.9322
\end{pmatrix}.
$$

We interpret these transition frequencies as conditional probabilities. For example, there is an 87.97% chance that squares that are currently in state $R$ (red squirrels only) will remain in state $R$; similarly, there is a 2.73% chance of squares that are currently occupied by both squirrel species, state $G$, will become occupied by neither species, state $B$, after the next time step. Since the data taken from the annual Forestry Commission survey is summarized as pairs of years, each time step represents a two-year period. The matrix form of the transition probabilities is convenient for calculations. Using matrix multiplication, we compute the two-time-step transition matrix as $T^2 = T \times T$, which is given by

$$
T^2 = \begin{pmatrix}
0.7799 & 0.0657 & 0.0930 & 0.0048 \\
0.0360 & 0.6432 & 0.0089 & 0.0250 \\
0.1733 & 0.0567 & 0.7834 & 0.0960 \\
0.0108 & 0.2344 & 0.1147 & 0.8742
\end{pmatrix}.
$$

The entries of this transition matrix are again interpreted as conditional probabilities. For instance, there is a 17.33% chance that squares currently occupied by only red squirrels, state $R$, will be occupied by both species, state $B$, in two time steps (four years).

Using the transition matrix $T$, it is possible to gain insight into the long-term behavior of the two species of squirrels. We compute the steady-state matrix $L$ for the two squirrel populations. The question of interest in the study of the squirrel populations is what happens to the distribution of the squirrel populations over a long period of time.
For our squirrel model, the steady-state matrix is approximately

\[ L \approx T^{200} = \begin{pmatrix}
0.1705 & 0.1705 & 0.1705 & 0.1705 \\
0.0560 & 0.0560 & 0.0560 & 0.0560 \\
0.3421 & 0.3421 & 0.3421 & 0.3421 \\
0.4314 & 0.4314 & 0.4314 & 0.4314 \\
\end{pmatrix}. \tag{1} \]

Thus the steady-state distribution is

\[ X = \begin{pmatrix}
0.1705 \\
0.0560 \\
0.3421 \\
0.4314 \\
\end{pmatrix}. \]

This result is interpreted as the long-term behavior of the squirrel populations in Great Britain as follows: 17.05% of the squares will be in state \( R \), containing only red squirrels. There will be 5.6% of the squares in state \( G \) containing only gray squirrels. There will be populations of both squirrels in 34.21% of the squares (state \( B \)), with the majority of the squares, 43.14%, being occupied by neither species of squirrels (state \( O \)).

If the assumptions made in this model are correct, the red squirrel is not currently in danger. In fact, it will have sole possession of more regions than the gray squirrel will have. In the long term, the gray squirrels do not drive the reds to extinction. Actually this analysis says nothing about population sizes, only about the number of regions controlled by each type of squirrel. While it seems plausible that if the red squirrel territory (number of regions) is declining, then the population is declining; the opposite may be true. A problem in the projects section asks you to perform this analysis for the two squirrel species in Scotland, where the red squirrel is still widely distributed.
2. Harvesting Scot Pines: An Absorbing Markov Chain Model

Scot pine trees (*Pinus sylvestris*) are grouped into classes according to girth. M. B. Usher, [4], studied the dynamics of a Scot pine forest in Corrour, Scotland. We wish to develop a model to study movement of Scot pine trees from one girth size to another and determine the average number of years to harvest.

We group the trees into six classes or states, labeled $i = 0, 1, \ldots, 5$. State 0 corresponds to the thinnest trees, and state 5 represents the thickest trees. Tree girth measurements are taken every six years. The girth classes are chosen so that, at each measurement, the trees either remain in their current class or gain sufficient girth to move into the next class, but never gain enough girth to skip a class. We also assume that all trees eventually reach full girth and, once in that state, are harvested. Thus state $i = 5$ is an absorbing state.

The assumption that a tree either remains in its current girth class or moves to the next larger girth class gives the following transition matrix for this Scot pine forest model.

$$
T = \begin{pmatrix}
p_0 & 0 & 0 & 0 & 0 & 0 \\
1 - p_0 & p_1 & 0 & 0 & 0 & 0 \\
0 & 1 - p_1 & p_2 & 0 & 0 & 0 \\
0 & 0 & 1 - p_2 & p_3 & 0 & 0 \\
0 & 0 & 0 & 1 - p_3 & p_4 & 0 \\
0 & 0 & 0 & 0 & 1 - p_4 & 1
\end{pmatrix}.
$$

(2)

This matrix is lower triangular. The last state, $i = 5$, being absorbing is reflected here by the entry in position 6, 6 being one. We need to determine the values of $p_0$, $p_1$, $p_2$, $p_3$, and $p_4$. Table 2 provides the data for a single census period and is taken from the paper by Usher [4] and is reprinted by permission of Blackwell Science Inc.

The value of $p_0$ is the proportion of trees that remain in state 0, which is $p_0 = 3,214/4,461 \approx 0.7205$, so that $1 - p_0 \approx 0.2795$. We similarly calculate $p_1 = 2,029/2,926 \approx$
Table 2: Scot Pine (*Pinus sylvestris*) Census Data from Corrour, Scotland.

0.6934, $p_2 \approx 0.7486$, $p_3 \approx 0.7703$, and $p_4 \approx 0.6296$. The transition matrix is then given as

$$T = \begin{pmatrix} 0.7205 & 0 & 0 & 0 & 0 & 0 \\ 0.2795 & 0.6934 & 0 & 0 & 0 & 0 \\ 0 & 0.3066 & 0.7486 & 0 & 0 & 0 \\ 0 & 0 & 0.2514 & 0.7703 & 0 & 0 \\ 0 & 0 & 0 & 0.2297 & 0.6296 & 0 \\ 0 & 0 & 0 & 0 & 0.3704 & 1 \end{pmatrix}. \quad (3)$$

The steady-state matrix is computed numerically with $T^n$ using a large value of $n$.

$$L \approx T^{200} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \quad (4)$$

This steady-state matrix should not come as a surprise. The matrix (4) shows that for each of the tree girth classes $i = 1, \ldots, 4$, members eventually reach full girth size, state $i = 5$, and are harvested. This is an obvious observation from the assumption that all trees survive to full girth size.
Using the block decomposition, we have $A_{5 \times 5}$ is the submatrix given by

$$A_{5 \times 5} = \begin{pmatrix}
0.7205 & 0 & 0 & 0 & 0 \\
0.2795 & 0.6934 & 0 & 0 & 0 \\
0 & 0.3066 & 0.7486 & 0 & 0 \\
0 & 0 & 0.2514 & 0.7703 & 0 \\
0 & 0 & 0 & 0.2297 & 0.6296
\end{pmatrix}.$$ 

The matrix $B_{1 \times 5}$ is the row

$$B_{1 \times 5} = (0 \ 0 \ 0 \ 0 \ 0.3704),$$

and $O_{5 \times 1}$ is the column vector of all zeros,

$$O_{5 \times 1} = \begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
0
\end{pmatrix}.$$ 

In this model, the matrix $I_{1 \times 1}$ is just the $1 \times 1$ identity matrix.

We may verify that in the Scot pine tree model,

$$B_{1 \times 5}(I_{5 \times 5} - A_{5 \times 5})^{-1} = (1 \ 1 \ 1 \ 1 \ 1).$$

For the Scot pine tree model, the fundamental matrix is

$$F = (I - A)^{-1} = \begin{pmatrix}
3.5778 & 0 & 0 & 0 & 0 \\
3.2149 & 3.2616 & 0 & 0 & 0 \\
3.9208 & 3.9777 & 3.9777 & 0 & 0 \\
4.2912 & 4.3535 & 4.3535 & 4.3535 & 0 \\
2.6611 & 2.6998 & 2.6998 & 2.6998 & 2.6998
\end{pmatrix}.$$ \hspace{1cm} (5)

Recalling the block form of the transition matrix, the position of the submatrix $A$ indicates that $i$ and $j$ have values $0, \ldots, 4$, so that $f_{00} = 3.5778$ is the average number of time steps
that a tree spends in the first girth state before moving up to the next girth state. Since each time step in this model represents a six-year period, we expect a tree to take an average of $3.58 \times 6 = 21.5$ years to move from the smallest girth size state 0 to state 1, the next largest state size. The sum of the entries for an individual column gives the average number of time steps to absorption. The Scot pine tree model predicts that on average, a tree of smallest girth size, state 0, will take about $(3.58 + 3.21 + 3.92 + 4.29 + 2.66) \times 6 \approx 106$ years to reach full girth (state 5).

Note that the lower-triangularity and the similarity of the entries in the fundamental matrix (5) reflect the assumptions of the Scot pine tree model.

We should expect that the average number of time steps required for a tree to move from one girth size to another would be independent of which state the tree began in, but once in a particular state, the average number of time steps depends upon that particular state. For example, we see in (5) that it takes about four time steps (24 years) for a tree to move from state 2 to state 3, regardless of which state the tree began in.

3. Hermit Crab Problem

The hermit crab (Pagurus longicarpus) does not have a hard protective shell to protect its body. It uses discarded shells to carry around as portable shelters. These empty shells are rare commodities in tide pools and are available only when their occupants die. In an experiment at Long Island Sound tide pools, an empty shell was dropped into the water to initiate a chain of vacancies. This experiment was repeated a large number of times as vacancies flowed from larger to generally smaller shells. The states in this experiment are the various shell sizes. Table 3.8 summarizes the number of moves between states.

There are seven states in this model. States 6 and 7 are absorbing states, which is reflected in the above table by the absence of moves from these states. State 6 represents the situation in which a crab without a shell occupies an empty shell. If a shell is not occupied during a 45-minute observation period, we assume that the shell is abandoned and label this
Table 3: Number of Moves between Shells in a Crab Vacancy Chain.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>19</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>20</td>
<td>11</td>
<td>10</td>
<td>4</td>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>26</td>
<td>26</td>
<td>6</td>
<td>24</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>22</td>
<td>2</td>
<td>30</td>
<td>59</td>
</tr>
</tbody>
</table>

as absorbing state 7.

1. Construct the state diagram.

2. Construct the transition matrix for this absorbing Markov chain.

3. Determine the steady-state matrix and the absorption probabilities.

4. Calculate the expected length of the crab vacancy chains.

This problem is adapted from work originally in the papers by Chase et al. [2], [6], and is discussed further in Beltrami’s modeling text [1].
Bibliography


