2 Transient behavior of Markov chains

In this exercise you learn to build Markov chain models, draw transition diagrams, and calculate transient distributions of Markov chains using the methods of matrix algebra. It is recommended to bring a laptop or a calculator to the exercise session to make it easier to calculate the numerical results of the matrix calculations arising in the exercises.

Classroom problems

2.1 *PageRank chain.* Google's PageRank algorithm is designed to rank web pages in terms of their centrality in the web graph. For a directed graph on a finite node set V, the PageRank of a node $v \in V$ is the limiting probability $\pi(v)$ of a Markov chain being in state v. The Markov chain has transition matrix

$$P(x,y) \;=\; c\,\frac{1}{n} + (1-c)\,\frac{G(x,y)}{\sum_{y' \in V} G(x,y')}, \qquad x,y \in V, \tag{1}$$

where $c \in [0,1]$ is called a damping factor, and G is the adjacency matrix of the graph so that

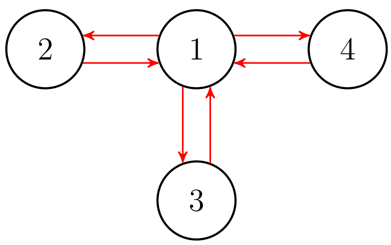
$$G(x,y) \;=\; \begin{cases} 1, & \text{if there is a link from } x \text{ to } y, \\ 0, & \text{else.} \end{cases}$$

Consider a directed graph on node set $V = \{1,2,\dots,n\}$ which contains the directed links $1 \to x$ and $x \to 1$ for all $x = 2,3,\dots,n$ but no other links.

(a) Draw the graph and write down its adjacency matrix G in the case $n = 4$.

Solution.

With $n = 4$ the graph is:



Directly from the definition of G we get

$$G = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

(b) Verify that the the n-by-n matrix defined by (1) is indeed a transition matrix. That is, verify that the entries of P are nonnegative and the row sums are equal to one.

Solution. The transition matrix of the PageRank model is defined by

$$P(x,y) = c\frac{1}{n} + (1-c)\frac{G(x,y)}{\sum_{y'\in V} G(x,y')},$$

where $|V| = n$ and G is the adjacency matrix of the graph.

i. Nonnegativity: For each $x, y \in V$ it holds that $0 \leq G(x,y) \leq \sum_{y'\in V} G(x,y')$. It is seen that $P(x,y)$ is a convex combination (weighted average) of the numbers $\dfrac{G(x,y)}{\sum_{y'\in V} G(x,y')}$ and $\dfrac{1}{n}$. Since these two numbers are between 0 and 1, so is $P(x,y)$.

ii. Row sums 1: for each $x \in V$ it holds that

$$\begin{aligned}\sum_{y\in V} P(x,y) &= c\frac{1}{n}\sum_{y\in V} 1 + (1-c)\frac{\sum_{y\in V} G(x,y)}{\sum_{y'\in V} G(x,y')} \\ &= c\frac{1}{n} \times n + (1-c) \times 1 \\ &= 1.\end{aligned}$$

(c) Write down the transition matrix P and draw its transition diagram in the case $n = 4$ for the values $c = 1/2$, $c = 0$ and $c = 1$ of the damping factor. How would you describe the behavior of the Markov chain in the case $c = 1$?

Solution. Denote $P_c = (P_c(i,j))_{1\leq i,j\leq 4}$, where

$$P_c(i,j) = \frac{1}{4}c + (1-c)\frac{G(i,j)}{\sum_{k=1}^4 G(i,k)}.$$

Hence,

$$P_{1/2} = \begin{bmatrix} 1/8 & 7/24 & 7/24 & 7/24 \\ 5/8 & 1/8 & 1/8 & 1/8 \\ 5/8 & 1/8 & 1/8 & 1/8 \\ 5/8 & 1/8 & 1/8 & 1/8 \end{bmatrix},$$

$$P_0 = \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad P_1 = \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}.$$

In the case $c = 0$ the transition diagram is the same as the graph drawn in (a). For $c \neq 0$ the transition diagram is a complete graph with four vertices.

Let $(X_t)_{t\geq 0}$ be the Markov chain corresponding to the transition matrix P_1. Now $\mathbb{P}(X_{t+1} = i \mid X_t = j) = 1/4$ for all i, j, t. In other words, $(X_t)_{t\geq 0}$ are independent and identically and uniformly distributed on the set $\{1, 2, 3, 4\}$.

(d) Suppose that $c \in (0, 1)$. What is the probability that a chain starting from state 1 is in state 1 after one time step?

Solution. A chain starting at vertex 1 is still at vertex 1 after one time step with probability

$$\begin{aligned} P_c(1,1) &= \frac{1}{n}c + (1-c)\frac{G(1,1)}{\sum_{k=1}^n G(1,k)} \\ &= \frac{1}{n}c + (1-c) \times 0 \\ &= \frac{1}{n}c. \end{aligned}$$

(e) What about after two time steps?

Solution. The probability in question is $P_c^2(1,1)$, or:

$$\begin{aligned} P_c^2(1,1) &= \sum_{s=1}^n P_c(1,s)P_c(s,1) \\ &= (P_c(1,1))^2 + \sum_{k=2}^n \left[\frac{1}{n}c + (1-c)\frac{G(1,s)}{\sum_{k=1}^n G(1,k)}\right]\left[\frac{1}{n}c + (1-c)\frac{G(s,1)}{\sum_{k=1}^n G(s,k)}\right] \\ &= (\frac{1}{n}c)^2 + \sum_{k=2}^n \left[\frac{1}{n}c + (1-c)\frac{1}{n-1}\right]\left[\frac{1}{n}c + (1-c)\frac{1}{1}\right] \\ &= \frac{c^2}{n^2} + \left[\frac{n-c}{n(n-1)}\right]\left[\frac{n+c-cn}{n}\right](n-1) \\ &= \frac{n-cn+c^2}{n}. \end{aligned}$$

Homework problems

2.2 *Espoo weather.* A simple model for October's weather in Espoo is a Markov chain with state space $S = \{1, 2, 3\}$ and transition matrix

$$P = \begin{bmatrix} 0.2 & 0.7 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.5 & 0.4 \end{bmatrix},$$

where 1 = 'rainy', 2 = 'cloudy', and 3 = 'sunny'.

(a) If it's cloudy tomorrow, then what is the probability that it's also cloudy the day after tomorrow? What about the day after that?

Solution.

Denote tomorrow's weather by $X_0 = 2$. The probability of 'cloudy' on the day after tomorrow (i.e., X_1) is

$$\mathbb{P}(X_1 = 2 \mid X_0 = 2) = P(2, 2) = 0.7.$$

On the other hand, the probability of 'cloudy' after the day after that equals

$$\mathbb{P}(X_2 = 2 \mid X_0 = 2) = P^2(2, 2) = 0.68,$$

which is seen from the matrix

$$P^2 = \begin{bmatrix} 0.19 & 0.68 & 0.13 \\ 0.19 & 0.68 & 0.13 \\ 0.16 & 0.62 & 0.22 \end{bmatrix}.$$

(b) If it's sunny next Sunday, then what is the probability that next Sunday is followed by at least four sunny days in a row?

Solution. In this case X_0 equals 3. The desired probability is

$$\begin{aligned} &\mathbb{P}(X_4 = 3, X_3 = 3, X_2 = 3, X_1 = 3 \mid X_0 = 3) \\ &\quad = \mathbb{P}(X_4 = 3 \mid X_3 = 3)\mathbb{P}(X_3 = 3 \mid X_2 = 3)\mathbb{P}(X_2 = 3 \mid X_1 = 3)\mathbb{P}(X_1 = 3 \mid X_0 = 3) \\ &\quad = (P(3, 3))^4 \\ &\quad = 0.0256 \end{aligned}$$

Additional information. The previous calculation can be justified as follows: denote

$$A_k = \text{"}X_k = 3\text{"}, \quad \text{for } k = 1, 2, 3, 4.$$

By the definition of conditional probability,

$$\begin{aligned} &\mathbb{P}(A_4 \cap A_3 \cap A_2 \cap A_1 \mid A_0) \\ &= \frac{\mathbb{P}(A_4 \cap A_3 \cap A_2 \cap A_1 \cap A_0)}{\mathbb{P}(A_0)} \\ &= \frac{\mathbb{P}(A_4 \cap A_3 \cap A_2 \cap A_1 \cap A_0)}{\mathbb{P}(A_3 \cap A_2 \cap A_1 \cap A_0)} \frac{\mathbb{P}(A_3 \cap A_2 \cap A_1 \cap A_0)}{\mathbb{P}(A_2 \cap A_1 \cap A_0)} \frac{\mathbb{P}(A_2 \cap A_1 \cap A_0)}{\mathbb{P}(A_1 \cap A_0)} \frac{\mathbb{P}(A_1 \cap A_0)}{\mathbb{P}(A_0)} \\ &= \mathbb{P}(A_4 \mid A_3 \cap A_2 \cap A_1 \cap A_0)\mathbb{P}(A_3 \mid A_2 \cap A_1 \cap A_0)\mathbb{P}(A_2 \mid A_1 \cap A_0)\mathbb{P}(A_1 \mid A_0). \end{aligned}$$

This formula is known as the product rule, or chain rule. By the Markov property this equals

$$\mathbb{P}(A_4 \mid A_3)\mathbb{P}(A_3 \mid A_2)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1 \mid A_0),$$

as claimed.

(c) Compute the weather distribution for the Saturday of week 44 (29 Oct – 2 Nov 2018), given that the Monday on that week is a cloudy day. Compute the weather distribution also for the corresponding Sunday, and find out for which of the weekend days rain is more probable.

Solution. Denote by $X_0 = 2$ the assumption that Monday is a cloudy day. The distribution of the initial state is then $\mu_0 = [0\ 1\ 0]$. The distributions of the states are

$$\mu_5 = [\mathbb{P}(X_5 = 1)\ \mathbb{P}(X_5 = 2)\ \mathbb{P}(X_5 = 3)],$$
$$\mu_6 = [\mathbb{P}(X_6 = 1)\ \mathbb{P}(X_6 = 2)\ \mathbb{P}(X_6 = 3)].$$

Using a computer we obtain

$$P^5 = \begin{bmatrix} 0.1858 & 0.6717 & 0.1425 \\ 0.1858 & 0.6717 & 0.1425 \\ 0.1850 & 0.6700 & 0.1449 \end{bmatrix},$$
$$P^6 = \begin{bmatrix} 0.1857 & 0.6715 & 0.1428 \\ 0.1857 & 0.6715 & 0.1428 \\ 0.1855 & 0.6710 & 0.1435 \end{bmatrix}.$$

From this we see that

$$\mu_5 = \mu_0 P^5 = [0.1858\ \ 0.6717\ \ 0.1425],$$
$$\mu_6 = \mu_0 P^6 = [0.1857\ \ 0.6715\ \ 0.1428].$$

Since $\mu_5(1) > \mu_6(1)$, Saturday is more likely to be rainy, although the probabilities are almost equal for Saturday and Sunday.

Additional information. In particular, all rows of P^6 are almost the same, so the weather model "forgets its initial state" in a week. The distribution $[0.1857\ \ 0.6715\ \ 0.1428]$ is the *limiting distribution* of the Markov chain, which does not depend on the initial state. This will be the next topic of the course.

2.3 *Genetic inheritance.* A trait of an individual is determined by a pair of genes which exist in two allelic versions: dominant allele `A` and recessive allele `a`. Hence each individual is characterized by one of the of three possible genotypes: `AA` (dominant homozygote), `Aa` (heterozygote[1]) and `aa` (recessive homozygote). The following is known about heredity:

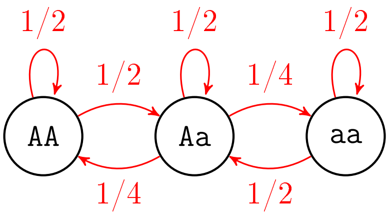
- Parents having genotypes `AA` and `Aa` produce children with genotypes `AA` and `Aa`, with equal probabilities.
- Parents having genotypes `aa` and `Aa` produce children of genotypes `aa` and `Aa`, with equal probabilities.
- Parents having identical genotypes `Aa` produce children having genotype `AA` with probability $\frac{1}{4}$, genotype `Aa` with probability $\frac{1}{2}$, and genotype `aa` with probability $\frac{1}{4}$.

Let us study the offspring of an individual with a dominant homozygous `AA` genotype for ten generations, when we assume that the initial individual and all its descendants produce offspring with heterozygotes (`Aa`-individuals).

(a) Write down the transition matrix of a Markov chain that models the genotypes of the offspring in future generations.

Solution. The transition matrix and the transition diagram of the Markov chain (with states $\{\mathtt{AA}, \mathtt{Aa}, \mathtt{aa}\}$) are:

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$



(b) Calculate the occupancy matrix M_{10} of the model.

Solution. Using a computer we get

$$\begin{aligned} M_{10} &= \sum_{s=0}^{9} P^s \\ &= I + P + P^2 + \cdots + P^9 \\ &\approx \begin{bmatrix} 3.75 & 4.5 & 1.75 \\ 2.25 & 5.5 & 2.25 \\ 1.75 & 4.5 & 3.75 \end{bmatrix}. \end{aligned}$$

[1] The combination `aA` is genetically equivalent to the combination `Aa` so we make no distinction between them.

Additional information. In practice the scripts might look, e.g., like this

```
% MATLAB
P = [1/2 1/2 0; 1/4 1/2 1/4; 0 1/2 1/2]
eye(3)+P+P^2+P^3+P^4+P^5+P^6+P^7+P^8+P^9
```

```
(* Mathematica *)
P = {{1/2, 1/2, 0}, {1/4, 1/2, 1/4}, {0, 1/2, 1/2}}
IdentityMatrix[3] + Sum[MatrixPower[P, k], {k, 1, 9}]
```

```
# R
library(expm)
P <- rbind(c(1/2,1/2,0),c(1/4,1/2,1/4),c(0,1/2,1/2))
res = diag(3)
for(k in 1:9) {res<-res+P%^%k}
res
```

(c) Using the occupancy matrix, find out the expected number of recessive homozygous ($\mathtt{aa}$) descendants of the dominant homozygote ($\mathtt{AA}$) in the first ten generations.

Solution. The occupancy of the state $\mathtt{aa}$ in the time interval $[0, 9]$ is

$$N_{\mathtt{aa}}(10) = \sum_{s=0}^{9} 1(X_s = \mathtt{aa}).$$

By the previous part we have

$$\mathbb{E}(N_{\mathtt{aa}}(10) \mid X_0 = \mathtt{AA}) = M_{10}(\mathtt{AA}, \mathtt{aa}) \approx 1.75.$$

Additional information. The model in this exercise has been simplified in the sense that we only study one gene and the "outside" genotype is assumed to always be $\mathtt{Aa}$. Nevertheless, heredity is modeled in genetics as random processes using similar ideas.