E Peltola

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, n is the number of nodes, 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 $V = \{1, 2, \ldots, n\}$ which contains the directed links $1 \to x$ and $x \to 1$ for all $x = 2, 3, \ldots, n$ but no other links.

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

(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.

(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 $c = 1$?

(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?

(e) What about after two time steps?

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?

(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?

(c) Compute the weather distribution for the Saturday of week 44 (31 Oct – 6 Nov 2022), 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.

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.

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

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

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