week08–markov

March 3, 2024

1 Math087 - Mathematical Modeling

Tufts University – Department of Math
George McNinch george.mcninch@tufts.edu
Fall 2020

1.1 Course material (Week 8): Stochastic matrices & Markov Chains

1.2 Probability, power iteration, and stochastic matrices

A vector $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T \in \mathbb{R}^n$ will be said to be a *probability vector* if all of its entries v_i satisfy $v_i \geq 0$ and if

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \cdot \mathbf{v} = \sum_{i=1}^n v_i = 1.$$

Let $A = \begin{bmatrix} a_{ij} \end{bmatrix} \in \mathbb{R}^{n \times n}$. We say that A is a *stochastic matrix* if $a_{ij} \geq 0$ for all i, j and if

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \cdot A = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix};$$

in words, A is a stochastic matrix if each column of A is a probability vector.

Notice that if $\mathbf{v}$ is a probability vector, and A is a stochastic matrix, then $A\mathbf{v}$ is again a probability vector.

Indeed, by the definitions we have

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \cdot A \cdot \mathbf{v} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \cdot \mathbf{v} = 1$$

As a consequence if A and B are stochastic $n \times n$ matrices, then also AB is stochastic. In particular, A^m is stochastic for all $m \geq 0$.

1.3 Eigenvalues of stochastic matrices

Proposition: Let A be a stochastic matrix.
a) A has an eigenvector with eigenvalue 1.
b) Let λ be any eigenvalue of a A. Then $|\lambda| \leq 1$.
c) If $\mathbf{w}$ is an eigenvector of A with eigenvalue λ satisfying $\lambda \neq 1$ then $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w} = 0$.

Sketch:

For **a)**, note that taking transposes and applying the definition, we find that

$$A^T \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T;$$

thus $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$ is an eigenvector for A^T with eigenvalue 1. Since the matrices A and A^T have the same characteristic polynomial and hence the same eigenvalues, the assertion **a)** now follows.

Since all entries a_{ij} of A satisfy $0 \le a_{ij} \le 1$, assertion **b)** is a consequence of Gershgorin's Theorem.

1.4 Proof of c):

On one hand, we have

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \lambda \mathbf{w} = \lambda \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w} \right)$$

On the other hand, since A is stochastic we have

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} A\mathbf{w} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w};$$

since $A\mathbf{w} = \lambda \mathbf{w}$ and since $\mathbf{w} \ne \mathbf{0}$, we conclude that

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w} = \lambda \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w}.$$

Since $\lambda \ne 1$ by assumption, this is only possible if $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{w} = 0$, as asserted.

1.5 Power iteration for stochastic matrices

Let A be a stochastic matrix, and *suppose* that the eigenvalue $\lambda = 1$ has multiplicity one. This means that the *1-eigenspace* has dimension 1.

More concretely, this means that $A - \mathbf{I_n}$ has rank $n - 1$.

Remark: If A has n distinct eigenvalues, then the each eigenspace has dimension 1.

We have the following:

1.6 Corollary

Suppose that the stochastic matrix A is diagonalizable, and that the *1-eigenspace* of A has dimension 1. Let $\mathbf{v}$ be an eigenvector for A with eigenvalue 1, and set $c = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{v}$. Then $\mathbf{w} = \dfrac{\mathbf{v}}{c}$ is a probability vector, and

$$A^m \to B \quad \text{as } m \to \infty$$

for a stochastic matrix B. Each column of B is equal to $\mathbf{w}$.

Sketch:

For $1 \le i \le n$, the i-th column of B may be computed as

$$\lim_{m\to\infty} A^m \mathbf{e}_i$$

where $\mathbf{e}_i$ is the i-th standard basis vector.

Let $\mathbf{v} = \mathbf{v}_1, \dots, \mathbf{v}_n$ be linearly independent eigenvectors for A.

When $j > 1$, the eigenvalue for $\mathbf{v}_j$ is < 1 by assumption, and it follows from the preceding results that $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \cdot \mathbf{v}_j = 0$ for $j > 1$.

Fix $1 \le i \le n$ and consider the expression

$$\mathbf{e}_i = \sum_{j=1}^{n} c_j \mathbf{v}_j.$$

Since $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{e}_i \ne 0$, it follows that $c_1 \ne 0$. Thus a result proved in the previous notebook shows that $\lim_{m\to\infty} A^m \mathbf{e}_i$ is a non-zero multiple of $\mathbf{w}$.

Since B is stochastic, each column of B is a probability vector, and must coincide with $\mathbf{w}$.

1.7 Markov Chains

Let's pause to recap our *finite-state machine* point-of-view.

We consider a system with a list of *states*. The system undergoes transitions, which we take to be given by probabilities.

We represent the system by a directed graph. Each state determines a node. A directed edge between two nodes $a \to b$ labeled with $p = p_{a,b}$ indicates that if the system is currently in state a, it will transform to state b with probability p.

Thus for each node a, the sum of the probabilities on the edges $a \to b$ must be 1:

$$\sum_{(a \to b)} p_{a,b} = 1$$

The resulting matrix $P = (p_{a,b})_{a,b}$ has the property that its column-sums are all equal to 1. Thus P is a *stochastic matrix*.

Let G be the directed graph attached to our probabilistic finite-state machine as before. We will refer to G as a *transition diagram*, and we call the *system* described by G a *Markov chain*.

1.8 Diagram properties

Let G be the transition diagram of a Markov chain.

Definition: G is *strongly connected* if for each pair of nodes a, b, there is sequence of directed edges $e_1, \dots, e_m$ connecting a to b.

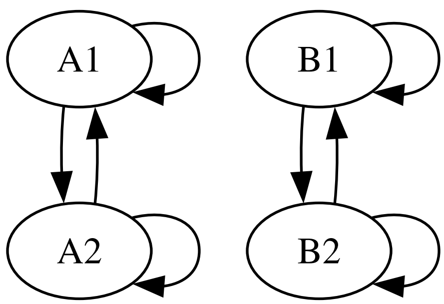
Remark: If P is the corresponding stochastic matrix, one often says that P is *irreducible* when the transition diagram G is *strongly connected*.

1.9 Example:

The following graph is not *strongly connected*.

[1]:
```
from graphviz import Digraph
from itertools import product
g = Digraph()

for i in ["A","B"]:
    for j in [1,2]:
        g.node(f"{i}{j}")
    for (j,k) in product([1,2],[1,2]):
        g.edge(f"{i}{j}",f"{i}{k}")

g
```

[1]:



1.10 Example:

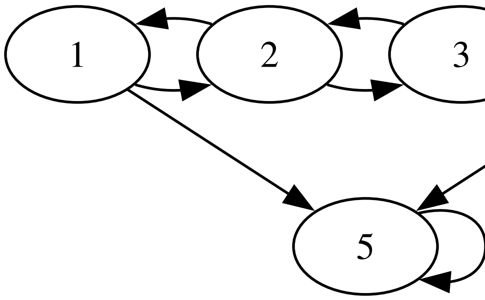
The following graph appears to be "connected" at least in some sense, but is not *strongly connected*.

Note that there is no path from the node 5 to the node 1, for example.

[2]:
```
h = Digraph()

I = [1,2,3,4]

with h.subgraph() as c:
    c.attr(rank="same")
    for i in I:
        c.node(f"{i}")
    for (j,k) in [(i,j) for (i,j) in product(I,I) if i == j+1 or i == j-1]:
        h.edge(f"{j}",f"{k}")

h.edge(f"1",f"5")
h.edge(f"4",f"5")
h.edge(f"5",f"5")
```

```
h
```

[2]:



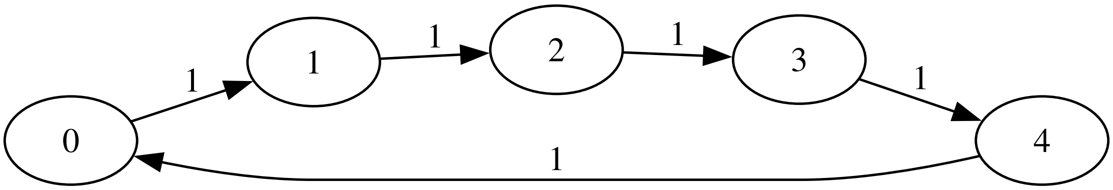
1.11 Cycles

A *cycle* of length n in a transition diagram is a sequence $e_1, \ldots, e_n$ of edges for which that initial node of e_1 is equal to the terminal node of e_n.

Here is an example of a cycle of length 5:

[3]:
```
import numpy as np

def cycle(n=5,labels=None):
    if labels==None:
        labels= n*[1]
    cyc = Digraph()
    cyc.attr(rankdir='LR')
    I = list(range(n))

    for i in I:
        cyc.node(f"{i}")

    for i in I:
        cyc.edge(f"{i}",f"{np.mod(i+1,n)}",f"{labels[i]}")

    return cyc

cycle()
```

[3]:



1.12 Aperiodic

Given a transition diagram G, consider all possible cycles in G.

A transition diagram is said to be *aperiodic* if no integer $n > 1$ divides the length of each cycle.

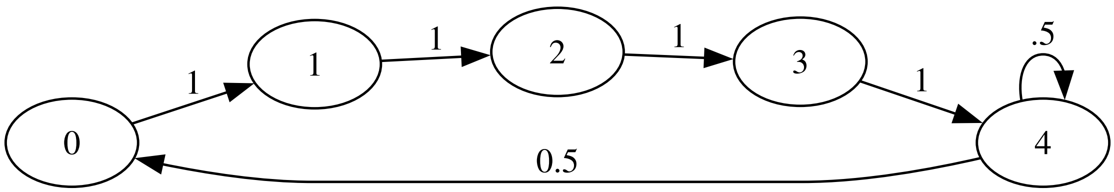
In other language, the diagram G is aperiodic if the greatest common divisor of the lengths of the cycles in G is equal to 1.

Example: The preceding graph G with 5 nodes is not *aperiodic* since every cycle has length a multiple of 5.

Example: The following graph *is* aperiodic, since it contains a cycle of length 1.

[4]:
```
acycle = cycle(labels=[1,1,1,1,.5])
acycle.edge("4","4",".5")
acycle
```

[4]:



1.13 Theorem: (Perron-Frobenius)

Let G be a transition diagram for a Markov chain, and suppose that G is strongly connected and aperiodic. Let P be the corresponding stochastic matrix. The multiplicity of the eigenvalue $\lambda = 1$ for P is 1 – i.e.

$$\dim \operatorname{Null}(P - I_n) = 1.$$

All other eigenvalues λ satisfy $|\lambda| < 1$.

There is a 1-eigenvector $\mathbf{v}$ which is a probability vector.

1.14 Corollary

a) $\lim\limits_{m\to\infty} P^m$ is a matrix for which each column is equal to $\mathbf{v}$.
b) If $\mathbf{w}$ is a vector for which $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \mathbf{v} > 0$, then $\lim\limits_{m\to\infty} P^m\mathbf{w}$ is a positive multiple of $\mathbf{v}$.

2 Financial market example

Recall this example from a previous *problem session.*

Consider the state of a financial market from week to week.

- by a *bull market* we mean a week of generally rising prices.

- by a *bear market* we mean a week of genreally declining prices.
- by a *recession* we mean a general slowdown of the economy.

Empirical observation shows for each of these three states what the probability of the state for the subsequent week, as follows:

	bull	*bear*	*recession*
followed by bull	0.90	0.15	0.25
followed by bear	0.075	0.80	0.25
followed by recession	0.025	0.05	0.50

In words, the first col indicates that if one has a bull market, then 90% of the time the next week is a bull market, 7.5% of the time the next week is a bear market, and 2.5% of the time the next week is in recession.

2.1 matrix

The matrix A describing the state transformations is a stochastic matrix.

[8]:
```
ones = np.ones(3)

A = np.array([[0.90 , 0.15 , 0.25],
              [0.075, 0.80 , 0.25],
              [0.025, 0.05 , 0.50]])

ones @ A
```

[8]:
```
array([1.00, 1.00, 1.00])
```

A has 3 distinct eigenvalues:

[10]:
```
##
e_vals,e_vecs = npl.eig(A)

e_vals
```

[10]:
```
array([1.00, 0.74, 0.46])
```

In particular, it follows that the 1-eigenspace of A has dimension 1.

A 1-eigenvector is given by

[12]:
```
v = e_vecs[:,0]
v
```

[12]:
```
array([0.89, 0.45, 0.09])
```

Rescaling `v` to make a probability vector, we indeed see that $A^m \to \begin{bmatrix} \mathbf{w} & \mathbf{w} & \mathbf{w} \end{bmatrix}$.

[29]:

```
float_formatter = "{:.4f}".format
np.set_printoptions(formatter={'float_kind':float_formatter})


w = (1/sum(v,0))*v

B=npl.matrix_power(A,200)

print(f"w = \n\n{w}\n\nA^200 = \n\n{B}")
```

```
w =

[0.6250 0.3125 0.0625]

A^200 =

[[0.6250 0.6250 0.6250]
 [0.3125 0.3125 0.3125]
 [0.0625 0.0625 0.0625]]
```

2.2 Interpretation:

Recall that A describes the state transitions for a financial market.

The interpretation here means that *in the long run*, there is a 62.5 % chance of a bull market, a 31.25 % chance of a bear market, and a 6.25% chance of a recession.