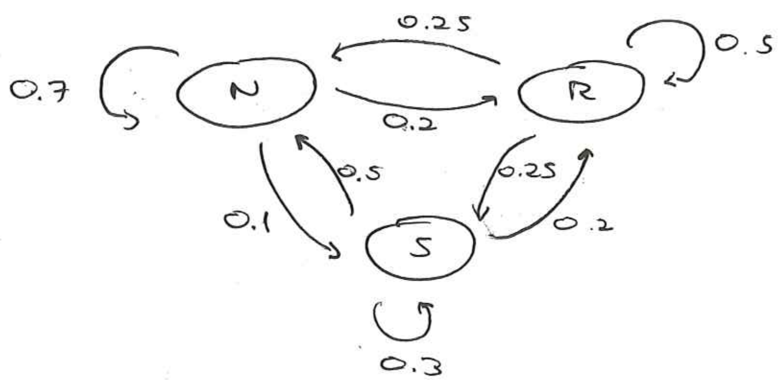
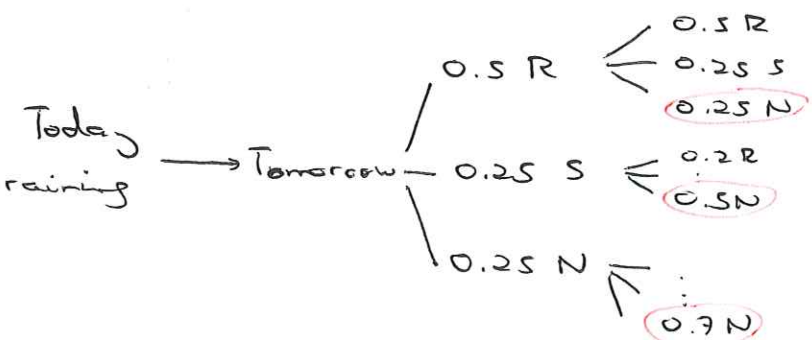
MATH 312 LECTURE 20

: Markov Chains.

Consider the following silly model for weather prediction:



If today is raining, what is the probability that the day after tomorrow is a nice day?



So after tomorrow,

$$p = (0.5)(0.25) + (0.25)(0.5) + (0.25)(0.7) = 0.425$$

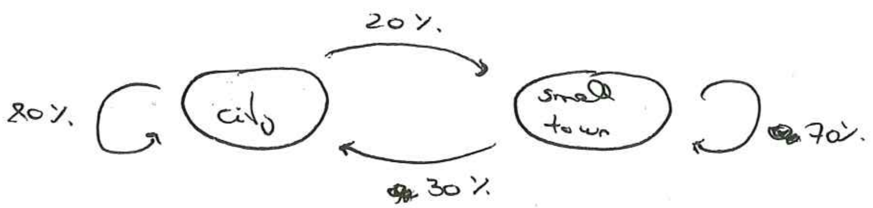
What is the probability that it will snow in a week, if today is a nice day?

Moreover, along the years, which fraction of days are nice, rainy or snowy?

→ The answers are not straightforward, but we will see they are easy to compute.

• Let's use a different simpler example:

(step = 1 year)



If today 50% of people live in cities, what is the distribution next year?

In cities: $(0.5)\cdot(0.8) + (0.5)(0.3) = 0.55 \to 55\%$.

In towns: $(0.5)\cdot(0.7) + (0.5)\cdot(0.2) = 0.45 \to 45\%$.

(same numbers)

And next one:

In cities: $(0.55)\cdot(0.8) + (0.45\cdot 0.3) = 0.575 \to 57'5\%$.

So in towns $\to 42'5\%$.

Create the "state" vector $\vec{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}$, where

$x_1(k) = \%$ in cities in year k

$x_2(k) = \%$ in town in year k

(of course, $x_1(k) + x_2(k) = 1 \quad \forall k$).

How to find $\vec{x}(k+1)$?

$$\left.\begin{aligned} x_1(k+1) &= x_1(k)\cdot(0.8) + x_2(k)\cdot(0.3) \\ x_2(k+1) &= x_1(k)\cdot(0.2) + x_2(k)(0.7) \end{aligned}\right\} \rightarrow \vec{x}(k+1) = \underbrace{\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}}_{\text{transition matrix}} \vec{x}(k)$$

Notice: columns add up to 1.

Q: What is the transition matrix for the weather model?

State vector: $\vec{x} = \begin{bmatrix} x_S \\ x_N \\ x_R \end{bmatrix}$,

$$\vec{x}(k+1) = P\,\vec{x}(k) \rightarrow P = \begin{bmatrix} 0.3 & 0.1 & 0.25 \\ 0.5 & 0.7 & 0.25 \\ 0.2 & 0.2 & 0.5 \end{bmatrix}$$

(P_{ij} = probability of going in one step from i to j)

Check previous result:

$$\underset{\substack{\uparrow \\ \text{probabilities} \\ \text{after two} \\ \text{days}}}{p} = P^2 \underset{\substack{\uparrow \\ \text{today} \\ \text{raining}}}{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} = P \begin{bmatrix} 0.25 \\ 0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} (0.3)(0.25)+(0.1)\cdot(0.25)+(0.5)\cdot(0.25) \\ (0.5)\cdot(0.25)+(0.7)\cdot(0.25)+(0.5)(0.25) \\ \cdots \end{bmatrix}$$

$\leftarrow$ probability of nice day after 2 days.

• Def: Markov (or stochastic) matrix

An $n \times n$ matrix is Markov if:

1) All columns add up to 1.

2) All entries are ≥ 0.

• The sequence of states $\{\vec{x}(1), \vec{x}(2), \dots, \vec{x}(k), \dots\}$ defined by a ~~a~~ Markov matrix is called a Markov chain (discrete)

Remark: key feature of this model: Every state only depends on the previous step: $\vec{x}(k+1) = A\vec{x}(k)$.

Example: 1) The evolution of a particle, defined through its position and velocity, is a Markov process (by Newton's law, if we know position and velocity now, we know the future).

ok, I'm considering classical mechanics.

2) Poker is not a Markov process: you use (or should use) the knowledge about which cards have been appearing from the start.

Some questions:

1) Each state $\vec{x}(k)$ describes the distribution at that moment

(% in cities, % in towns, and so on).

So $\vec{x}(k+1)$ needs to be a % distribution too, i.e, the components should add up to 1.

Is it true? That is, for A Markov matrix, does $A\vec{x}$ add up to 1 if $\vec{x}$ does?

Let's see that yes:

Notice that the addition of the components of a vector is given by

$$[1 \dots 1]\vec{x} = [1 \dots 1]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 + \dots + x_n = 1 .$$

assumption (%...)

But then,

$$[1 \dots 1] A\vec{x} = [1 \dots 1]\left[\vec{a}_1 \middle| \dots \middle| \vec{a}_n\right]\vec{x} = [1 \dots 1]\vec{x} = x_1 + \dots + x_n = 1 .$$

Markov matrix $\Rightarrow$ $\begin{cases} [1 \dots 1]\vec{a}_1 = \text{addition components first column} = 1 \\ [1 \dots 1]\vec{a}_2 = \text{" " second column} = 1 \\ \vdots \quad \text{All cols. of A add up to 1!} \end{cases}$

• <u>Def</u>: Steady state

Let $\{\vec{x}_0, \dots, \vec{x}_k, \dots\}$ be a Markov chain. We called steady state to the limit

$$\vec{x}_\infty = \lim_{k\to\infty} \vec{x}(k) \qquad \text{(if it exists)}.$$

<u>Remark</u>: Notice that if $\vec{x}_\infty$ exists, then

$$\vec{x}_\infty = \lim_{k\to\infty} \vec{x}(k) = \lim_{k\to\infty} A\vec{x}(k-1) = A \lim_{k\to\infty} \vec{x}(k-1) = A\vec{x}_\infty .$$

That is, $\| A\vec{x}_\infty = \vec{x}_\infty \|$ Eigenvector of A with eigenvalue 1 .
(with cols. adding up to 1)

2) (Question) Does a Markov chain always have a steady state?

→ General answer: <u>No</u> .

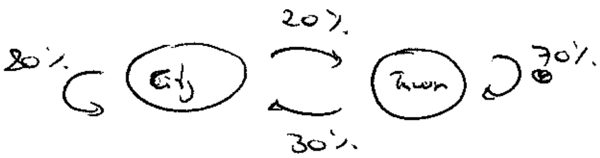
3) If the steady exists, does it depend on the initial configuration?

→ General answer: <u>Yes</u>

However, many systems <u>do</u> have a steady state and it <u>is</u> independent of the initial state.

Example: Find the steady distribution of people in cities and towns, if the initial configuration is

$$\vec{x}(0) = \begin{bmatrix} a \\ 1-a \end{bmatrix} \quad (100a\,\% \text{ in cities, } 100(1-a)\% \text{ in towns}).$$



$$P = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

Sol:

$$\vec{x}_\infty = \lim_{k\to\infty} P^k \, \vec{x}(0) \quad \to \quad P^k = S D^k S^{-1} \quad (\text{if } P \text{ diagonalizable}).$$

Eigenvalues: $P-\lambda I \rightsquigarrow (0.8-\lambda)(0.7-\lambda)-(0.2)(0.3)=0 \Leftrightarrow$

$\Leftrightarrow (8-10\lambda)(7-10\lambda)-6=0$ $\qquad \lambda_1 = 1$

$\Leftrightarrow 56+100\lambda^2-150\lambda-6 = 10\lambda^2-15\lambda+5=0 \to \lambda_2 = 1/2$

Eigvectors: $P - I = \begin{bmatrix} -0.2 & 0.3 \\ 0.2 & -0.3 \end{bmatrix} \to \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} \rightsquigarrow \vec{u}_1 = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix}$

not needed; just so that cols. add to 1.

$$P - \tfrac{1}{2} I = \begin{bmatrix} 0.3 & 0.3 \\ 0.2 & 0.2 \end{bmatrix} \to \vec{u}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

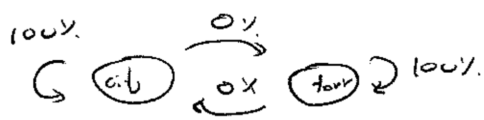
$$P = \begin{bmatrix} 0.6 & 1 \\ 0.4 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ -0.4 & 0.6 \end{bmatrix} (-1)$$

$$\vec{x}_\infty = P^\infty \vec{x}(0) = \begin{bmatrix}0.6 & 1\\ 0.4 & -1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 0.4 & -0.6\end{bmatrix}\begin{bmatrix}a\\ 1-a\end{bmatrix} =$$

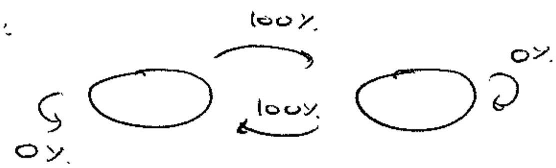
$$= \begin{bmatrix}0.6 & 0\\ 0.4 & 0\end{bmatrix}\begin{bmatrix}a+1-a\\ 0.4a-0.6+0.6a\end{bmatrix} = \begin{bmatrix}0.6\\ 0.4\end{bmatrix} \to \begin{matrix}\text{60\% cities}\\ \text{40\% towns}\end{matrix}$$

__Remark:__ $\lambda_1 = 1,\ \lambda_2 = \frac{1}{2} \to |\lambda_2| < 1.$

__Example 2:__ $P = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \to$ obviously $\vec{x}_\infty = \vec{x}(0) \to$ so it depends on initial state



__Example 3:__



$P = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \to \lambda^2 - 1 = 0$

$\hookrightarrow \lambda_1 = 1,\ \lambda_2 = -1.$

There is not steady state!

$\vec{x}(0) = \begin{bmatrix}a\\ 1-a\end{bmatrix} \to \vec{x}(1) = \begin{bmatrix}1-a\\ a\end{bmatrix} \to \vec{x}(2) = \begin{bmatrix}a\\ 1-a\end{bmatrix} \to \dots$ (unless $a = 0.5$).

Theorem: If A is a Markov matrix with

1) largest eigenvalue $\lambda = 1$ (non repeated)
2) all other eigenvalues $|\lambda| < 1$,

then, for any $\vec{x}(0)$ with ~~positive~~ nonnegative entries and components adding up to 1 (i.e. initial configuration),

we have that

$\to$ $\vec{x}_\infty = \lim_{k\to\infty} A^k \vec{x}(0)$ is an eigenvector of A for $\lambda = 1$.

$\to$ $\vec{x}_\infty$ is independent of the choice $\vec{x}(0)$.

Idea of the "proof": Assume A is diagonalizable, so there is a basis of eigenvectors $\{\vec{u}_1, \dots, \vec{u}_n\}$.

Write $\vec{x}(0) = c_1\vec{u}_1 + \dots + c_n\vec{u}_n$, and recall that components of $\vec{x}(0)$ add up to one.

Then,

$$\vec{x}(k) = A^k\vec{x}(0) = c_1 A^k \vec{u}_1 + c_2 A^k \vec{u}_2 + \dots + c_n A^k \vec{u}_n =$$

$$= c_1\vec{u}_1 + c_2\lambda_2^k \vec{u}_2 + \dots + c_n \lambda_n^k \vec{u}_n \Rightarrow$$

$$\Rightarrow \vec{x}_\infty = \lim_{k\to\infty} A^k \vec{x}(0) = c_1\vec{u}_1 + \vec{0} + \dots + \vec{0} = c_1\vec{u}_1.$$

It might seem that $\vec{x}_\infty$ depends on $\vec{x}_0$ through that c_1.

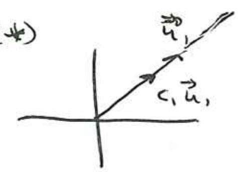
But recall that if $\vec{x}(0)$'s components add up to one, so does $\vec{x}(1), \vec{x}(2), \ldots$ and therefore $\vec{x}_\infty$. Thus,

$\vec{u}_1$ fixed by A (is an eigenvector for $\lambda=1$)

length of $c_1\vec{u}_1$ fixed since its components add up to 1 $\Big\} \rightarrow$

(*)

$\rightarrow$ No dependence on $\vec{x}(0)$

(*) 

Note that $\vec{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix}$ gives the slope of the line $m = \frac{u_{12}}{u_{11}}$ (fixed by A).

Then, any vector $\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ on that line is written as $\begin{bmatrix} \alpha_1 \\ m\alpha_1 \end{bmatrix}$

So if components add up to $1 \rightarrow \alpha_1 + m\alpha_1 = 1 \rightarrow \alpha_1 = \frac{1}{1+m}$ (fixed by A.)

- Which matrices satisfy the above theorem? $\rightarrow$ Sufficient condition: A Markov with strictly positive entries.

$\Downarrow$

(convergence of next theorem)

$\hookrightarrow$ Goal of the rest of chapter. // (pagerank example).

- Def: Spectral radius of a matrix, $\rho(A)$.

$$\rho(A) = \max\{|\lambda| : \lambda \text{ eigenvalue of } A\}.$$

Example: $\rho\left(\begin{bmatrix} 3+2i & 0 \\ 6 & 1 \end{bmatrix}\right) = \max\{|3+2i|, 1\} = \sqrt{9+4}$

$$\rho\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = 0$$

Remarks: 1) If A diagonalizable and $\rho(A)=0$, then $A=0$.
($\Leftarrow D=0$!)

2) $\rho(\frac{1}{\alpha}A)=\frac{1}{\alpha}\rho(A)$

Theorem: Perron-Frobenius

Let A be $n\times n$ with all entries strictly positive. Then,

1) $\rho(A)$ is a non-repeated eigenvalue of A.

2) All other eigenvalues satisfy $|\lambda| < \rho(A)$.
(strictly)

3) $\rho(A)$ has an eigenvector with all entries positive.

4) No other eigenvalue of A has a strictly positive eigenvector.
(meaning positive entries).