Let Monday be time t so Thursday is day $t+3$.

Then $\underset{\sim}{x}(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and

$$\underset{\sim}{x}(t+3) = P\,\underset{\sim}{x}(t+2) = P\,P\,\underset{\sim}{x}(t+1) = P\,P\,P\,\underset{\sim}{x}(t)$$

$$= P^3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 1/4 \end{pmatrix}$$

$\Rightarrow$ prob of pizza on Thursday is

$x_2(t+3) = 1/4$.

Definition 6.3.2. *A square matrix P is **stochastic** if each entry of P lies in $[0,1]$, and the entries in each column of P sum to 1.*

Definition 6.3.3. *A random process whose distribution satisfies*

$$\mathbf{x}(t+1) = P\,\mathbf{x}(t) \tag{6.2}$$

*with P a stochastic matrix is called a **Markov chain**. The matrix P is the **transition matrix** for the chain.*

$\boxed{(P)_{ji}}$ gives the probability of moving from state i to state j in one step.

If a Markov chain has initial distribution $\mathbf{x}(0)$, then by iterating (6.2) we obtain

$$\boxed{\mathbf{x}(t) = P^t\,\mathbf{x}(0).} \tag{6.3}$$

(6.3) holds by def^{n} for $t=1$. If it holds for some $t \geqslant 1$, then $\underset{\sim}{x}(t+1) = P\underset{\sim}{x}(t)$

$$= P\,P^t\,\underset{\sim}{x}(0)$$

$$= P^{t+1}\,\underset{\sim}{x}(0)$$

So (6.3) follows by induction.

E.g. 6.3.4. *Consider the lunching mathematician of E.g. 6.3.1. By diagonalizing the transition matrix P, find an exact expression for P^t for general t. What is $\lim_{t\to\infty} P^t$?*

Using (6.3), find the distribution of the mathematician's lunch choices on day t, given that $\mathbf{x}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. What is $\lim_{t\to\infty} \mathbf{x}(t)$?

How does the answer change if $\mathbf{x}(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$?

What about $\mathbf{x}(0) = \begin{bmatrix} p \\ 1-p \end{bmatrix}$ for arbitrary $p \in [0,1]$?

$P = QDQ^{-1}$ where $D = \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$, $Q = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \Rightarrow$

$$P^t = QD^tQ^{-1} = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (-\frac{1}{2})^t \end{pmatrix} \frac{1}{3} \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 2 + (-\frac{1}{2})^t & 2 - 2(-\frac{1}{2})^t \\ 1 - (-\frac{1}{2})^t & 1 + 2(-\frac{1}{2})^t \end{bmatrix} \xrightarrow[t\to\infty]{} \begin{pmatrix} \frac{2}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$\underset{\sim}{x}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$: $\underset{\sim}{x}(t) = P^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} \xrightarrow[t\to\infty]{} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$

$\underset{\sim}{x}(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$: $\underset{\sim}{x}(t) = P^t \begin{pmatrix} 0 \\ 1 \end{pmatrix} \xrightarrow[t\to\infty]{} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$

$\underset{\sim}{x}(0) = \begin{pmatrix} p \\ 1-p \end{pmatrix}$: $\underset{\sim}{x}(t) = P^t \begin{pmatrix} p \\ 1-p \end{pmatrix} \xrightarrow[t\to\infty]{} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$

Definition 6.3.5. *A stochastic matrix is* ***regular*** *if some finite power of P has all entries positive.*

Theorem 6.3.6. *If P is a regular stochastic matrix, then:* Perron-Frobenius

(a) There is a unique probability vector $\mathbf{q}$ such that $\boxed{P\mathbf{q} = \mathbf{q}}$

(b) For any initial distribution $\mathbf{x}(0) = \boldsymbol{\mu}$, the sequence $\boldsymbol{\mu}, P\boldsymbol{\mu}, P^2\boldsymbol{\mu}, \ldots$ converges to $\mathbf{q}$.

(Markov chain Convergence Th^{m})

The limiting distribution $\mathbf{q}$ is the **stationary distribution** of P.

Theorem 6.3.6 tells us that in order to find the limiting distribution of a regular Markov chain, we need only find the eigenvector corresponding to the eigenvalue 1, then normalize it appropriately.

E.g. 6.3.7. *Consider the lunching mathematician of E.g. 6.3.1. Use Theorem 6.3.6 to determine the stationary distribution.*

If $\underset{\sim}{x}(0) = \underset{\sim}{q}$ then $\underset{\sim}{x}(1) = P\underset{\sim}{x}(0) = \underset{\sim}{q}$

$\underset{\sim}{x}(2) = P\underset{\sim}{x}(1) = \underset{\sim}{q}$

$\vdots$

$\underset{\sim}{x}(t) = \underset{\sim}{q}$

If P is regular, the eigenspace for $\lambda = 1$ is $\text{null}(I - P) = \text{span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$

The stationary dist^{n} of P is $\underset{\sim}{q} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}$

E.g. 6.3.8. *Consider the stochastic matrix* $P = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/3 \\ 1/4 & 0 & 2/3 \end{bmatrix}$. *Verify that* P *is regular, and find its stationary distribution.*

$$P^2 = \frac{1}{16\times 9}\begin{bmatrix} 54 & 72 & 24 \\ 48 & 54 & 56 \\ 42 & 18 & 64 \end{bmatrix}$$

P^2 has positive entries $\Rightarrow$ P is regular.

So stationary distⁿ of P is the unique probability vector in null$(I-P)$

$$I-P \sim \begin{pmatrix} 1 & -1 & 0 \\ 0 & 3 & -4 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \text{null}(I-P) = \text{span}\left\{\begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}\right\}$$

$$\Rightarrow \quad \underset{\sim}{q} = \left(\tfrac{4}{11}, \tfrac{4}{11}, \tfrac{3}{11}\right).$$

6.4 Application: Search Engines (A&R §10.20)

The PageRank algorithm, invented by Google founder Larry Page, is used by Google to rank web pages.

Consider a subset of the world wide web consisting of n web sites, and let A denote the adjacency matrix of the corresponding directed graph

$$(A)_{ij} = \begin{cases} 1, & \text{site } i \text{ links to site } j, \\ 0, & \text{otherwise.} \end{cases}$$

PageRank models the movement of random web surfer as a Markov chain with transition matrix

$$(P)_{ji} = \frac{1-\delta}{n} + \delta \, \frac{(A)_{ij}}{\sum_{k=1}^{n} (A)_{ik}},$$

where $\delta \in (0,1)$ is a **damping factor**, usually taken to be $\delta = 0.85$.

In words, with probability δ, the surfer moves to a random site to which its current site links, and with probability $1-\delta$ the surfer moves to any random site at all.

PageRank ranks sites according to the stationary distribution of this Markov chain.

Note that $(P)_{ji} > 0$ for all i, j, so that P is regular. Theorem 6.3.6 therefore guarantees there exists a unique stationary distribution.

E.g. 6.4.1. *Consider a toy web graph with adjacency matrix*

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

Use PageRank with $\delta = 0.85$ *to rank the sites.*

In principle, one could find the eigenspace of P corresponding to eigenvalue 1, & choose the appropriately normalized element to obtain the exact stationary dist^{n}, $\underset{\sim}{q}$.
In practice, this is computationally infeasible for large networks.

Instead, take an arbitrary initial dist^{n} such as $\mu = (\frac{1}{n}, \dots, \frac{1}{n})$ & use the fact that $P^t \underset{\sim}{\mu} \xrightarrow[t\to\infty]{} \underset{\sim}{q}$ to approximate $\underset{\sim}{q}$.

$P^5 \underset{\sim}{\mu} = (0.1891, 0.348, 0.355, 0.03518, 0.025, 0.0477)$

which gives a ranking: 3, 2, 1, 6, 4, 5

For comparison, $\underset{\sim}{q}$ differs from $P^5 \underset{\sim}{\mu}$ in 3rd decimal place, & gives same ranking.

6.5 Application: Differential equations (A&R §5.4)

OMITTED

Recall that

$$\frac{dy}{dt} = a\,y(t) \tag{6.4}$$

is a **differential equation**.

In particular, it is a first-order, linear, homogeneous differential equation.

Recall also that (6.4) has the general solution $y(t) = c\,e^{a\,t}$.

Suppose that we want to solve a system of n coupled, first-order, linear, homogeneous differential equations

$$\begin{array}{rcccccccc} y_1' &=& a_{11}\,y_1 &+& a_{12}\,y_2 &+& \ldots &+& a_{1n}\,y_n \\ y_2' &=& a_{21}\,y_1 &+& a_{22}\,y_2 &+& \ldots &+& a_{2n}\,y_n \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ y_n' &=& a_{n1}\,y_1 &+& a_{n2}\,y_2 &+& \ldots &+& a_{nn}\,y_n \end{array} \tag{6.5}$$

How can we do it?

As a first observation, note that we can re-express this system as

$$\mathbf{y}' = A\mathbf{y}$$

where $A = [a_{ij}]$ is the coefficient matrix of the right-hand side of (6.5).

OMITTED

E.g. 6.5.1. *Find the general solution to the system*

$$\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

OMITTED

Q: How can we solve systems for which the coefficient matrix is not diagonal?
A: For systems with diagonalizable coefficient matrices, we can simply replace the system with an equivalent diagonal system.

The following method works quite generally for any system of n first order, linear, homogeneous differential equations, provided the coefficient matrix of its matrix form is diagonalizable.

Algorithm 6.5.2.

(i) *Write the system in matrix form* $\mathbf{y}' = A\mathbf{y}$

(ii) *Find a matrix* P *that diagonalizes* A

(iii) *Make the change of variables* $\mathbf{y} = P\mathbf{u}$

(iv) *Solve* $\mathbf{u}' = D\mathbf{u}$ *where* $D = P^{-1}AP$

(v) *Determine* $\mathbf{y}$ *from* $\mathbf{y} = P\mathbf{u}$

OMITTED

E.g. 6.5.3. *Find the general solution to the system*

$$\frac{dy_1}{dt} = y_1 + y_2,$$
$$\frac{dy_2}{dt} = 4y_1 - 2y_2.$$

Find the solution satisfying $y_1(0) = 1$, $y_2(0) = 6$.

6.6 The Power Method (A&R §9.2)

OMITTED

Solving the characteristic equation for large matrices is computationally difficult. Numerical methods for estimating eigenvalues and eigenvectors are therefore of significant practical importance.

Definition 6.6.1. *If the distinct eigenvalues of a matrix A are λ_1, λ_2, ..., λ_p and if $|\lambda_1| > |\lambda_2|, |\lambda_3|, \ldots, |\lambda_p|$ then λ_1 is a* ***dominant eigenvalue*** *of A, and any eigenvector corresponding to λ_1 is a* ***dominant eigenvector*** *of A.*

Theorem 6.6.2. *Let A be a symmetric $n \times n$ matrix with a positive dominant eigenvalue λ. If $\mathbf{x}_0 \in \mathbb{R}^n$ is a unit vector that is not orthogonal to the eigenspace of λ, then the sequence*

$$\mathbf{x}_0,\ \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|},\ \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\|A\mathbf{x}_1\|},\ \ldots,\ \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\|A\mathbf{x}_{k-1}\|},\ \ldots$$

converges to a unit dominant eigenvector, and the sequence

$$A\mathbf{x}_1 \cdot \mathbf{x}_1,\ A\mathbf{x}_2 \cdot \mathbf{x}_2,\ \ldots,\ A\mathbf{x}_k \cdot \mathbf{x}_k, \ldots$$

converges to the dominant eigenvalue λ.

By truncating these sequences one can obtain numerical approximations to the dominant eigenvalue and eigenvectors.

E.g. 6.6.3. *Let*

OMITTED

$$A = \begin{bmatrix} 5 & -1 \\ -1 & -1 \end{bmatrix}, \qquad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Apply the power method to A, starting at $\mathbf{x}_0$ and stopping at $\mathbf{x}_4$. Compare the resulting approximations to the exact values of the dominant eigenvalue and corresponding unit eigenvector.

6.7 Complex vector spaces (A&R §5.3)

OMITTED

E.g. 6.7.1. *Find the eigenvalues of*

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

As this e.g. shows, even real matrices can have complex eigenvalues. This motivates the study of complex vector spaces.

Theorem 6.7.2. *If λ is an eigenvalue of a real $n \times n$ matrix A, and if $\mathbf{x} \in \mathbb{C}^n$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of A, and $\overline{\mathbf{x}}$ is a corresponding eigenvector.*

Theorem 6.7.3. *Real symmetric matrices have real eigenvalues.*

Remark 6.7.4. *Complex vector spaces play a fundamental role in quantum mechanics. In QM, "states" of physical systems are represented by vectors in complex vector spaces, "observables" (such as energy, momentum, etc.) are represented by linear operators acting on these complex vector state spaces, and the possible (quantized) values that can be obtained by these observables are given by the eigenvalues of their linear operators. In this context, linear operators which are guaranteed to have real eigenvalues play a distinguished role.*