Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 16

16 March 2012

The story of the film so far...

- We are developing a language to study systems with a non-deterministic time evolution.

The story of the film so far...

- We are developing a language to study systems with a non-deterministic time evolution.
- More precisely, a **stochastic process** is a collection of random variables $\{X_t\}$ indexed by “time” taking values in a **state space** $\mathcal{S}$: X_t is the state of the system at time t.

The story of the film so far...

- We are developing a language to study systems with a non-deterministic time evolution.
- More precisely, a **stochastic process** is a collection of random variables $\{X_t\}$ indexed by “time” taking values in a **state space** $\mathcal{S}$: X_t is the state of the system at time t.
- A **Markov chain** $\{X_0, X_1, X_2, \dots\}$ is a discrete-time stochastic process with countable $\mathcal{S}$ satisfying the **Markov property**:

$$\mathbb{P}(X_{n+1} = s_{n+1} \mid X_0 = s_0, \dots, X_n = s_n) \\ = \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n)$$

The story of the film so far...

- We are developing a language to study systems with a non-deterministic time evolution.
- More precisely, a **stochastic process** is a collection of random variables $\{X_t\}$ indexed by "time" taking values in a **state space** $\mathcal{S}$: X_t is the state of the system at time t.
- A **Markov chain** $\{X_0, X_1, X_2, \ldots\}$ is a discrete-time stochastic process with countable $\mathcal{S}$ satisfying the **Markov property**:

$$\mathbb{P}(X_{n+1} = s_{n+1} \mid X_0 = s_0, \ldots, X_n = s_n) = \mathbb{P}(X_{n+1} = s_{n+1} \mid X_n = s_n)$$

- Markov chains are described by **stochastic** matrices $\mathbf{P}$ with $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ for all n, such that

$$p_{ij} \geqslant 0 \qquad \text{and} \qquad \sum_j p_{ij} = 1$$

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $P(m, m+n)$ be the **n-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix, $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix, $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m, and we will show that $\mathbf{P}(m, m+n) = \mathbf{P}^n$ for all m.

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix, $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m, and we will show that $\mathbf{P}(m, m+n) = \mathbf{P}^n$ for all m.
This will follow from the **Chapman–Kolmogorov formula**

$$\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$$

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix, $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m, and we will show that $\mathbf{P}(m, m+n) = \mathbf{P}^n$ for all m.
This will follow from the **Chapman–Kolmogorov formula**

$$\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$$

or in terms of probabilities

$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r)$$

n-step transition matrix

Consider a (temporally) homogeneneous Markov chain and let $\mathbf{P}(m, m+n)$ be the n**-step transition matrix** with entries

$$p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$$

It is again an stochastic matrix, $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m, and we will show that $\mathbf{P}(m, m+n) = \mathbf{P}^n$ for all m.
This will follow from the **Chapman–Kolmogorov formula**

$$\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$$

or in terms of probabilities

$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r)$$

The proof is not hard and uses the Markov property and some basic facts about probability.

Proof of the Chapman–Kolmogorov formula

By the partition rule,

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j, X_{m+n} = k \mid X_m = i)$$

Proof of the Chapman–Kolmogorov formula

By the partition rule,

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j,\, X_{m+n} = k \mid X_m = i)$$

Since $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid B \cap C)\mathbb{P}(B \mid C)$,

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j \mid X_{m+n} = k,\, X_m = i) \times \mathbb{P}(X_{m+n} = k \mid X_m = i)$$

Proof of the Chapman–Kolmogorov formula

By the partition rule,

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j,\, X_{m+n} = k \mid X_m = i)$$

Since $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid B \cap C)\mathbb{P}(B \mid C)$,

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j \mid X_{m+n} = k,\, X_m = i) \times \mathbb{P}(X_{m+n} = k \mid X_m = i)$$

and by the Markov property

$$\mathbb{P}(X_{m+n+r} = j \mid X_m = i) = \sum_k \mathbb{P}(X_{m+n+r} = j \mid X_{m+n} = k) \times \mathbb{P}(X_{m+n} = k \mid X_m = i)$$

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n.

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity).

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity). Now for the induction step, suppose that $\mathbf{P}(m, m+k) = \mathbf{P}^k$ for all m and for all $k < n$.

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity). Now for the induction step, suppose that $\mathbf{P}(m, m+k) = \mathbf{P}^k$ for all m and for all $k < n$. Then by the Chapman–Kolmogorov formula for $(m, n-1, 1)$,

$$\mathbf{P}(m, m+n) = \mathbf{P}(m, m+n-1)\mathbf{P}(m+n-1, m+n)$$

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity). Now for the induction step, suppose that $\mathbf{P}(m, m+k) = \mathbf{P}^k$ for all m and for all $k < n$. Then by the Chapman–Kolmogorov formula for $(m, n-1, 1)$,

$$\mathbf{P}(m, m+n) = \mathbf{P}(m, m+n-1)\mathbf{P}(m+n-1, m+n)$$

but $\mathbf{P}(m, m+n-1) = \mathbf{P}^{n-1}$ by the induction hypothesis

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity). Now for the induction step, suppose that $\mathbf{P}(m, m+k) = \mathbf{P}^k$ for all m and for all $k < n$. Then by the Chapman–Kolmogorov formula for $(m, n-1, 1)$,

$$\mathbf{P}(m, m+n) = \mathbf{P}(m, m+n-1)\mathbf{P}(m+n-1, m+n)$$

but $\mathbf{P}(m, m+n-1) = \mathbf{P}^{n-1}$ by the induction hypothesis, and $\mathbf{P}(m+n-1, m+n) = \mathbf{P}$

Corollary (of Chapman–Kolmogorov formula)

For all m, $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof.

By induction on n. For $n = 1$, we have that $\mathbf{P}(m, m+1) = \mathbf{P}$ for all m (temporal homogeneity). Now for the induction step, suppose that $\mathbf{P}(m, m+k) = \mathbf{P}^k$ for all m and for all $k < n$. Then by the Chapman–Kolmogorov formula for $(m, n-1, 1)$,

$$\mathbf{P}(m, m+n) = \mathbf{P}(m, m+n-1)\mathbf{P}(m+n-1, m+n)$$

but $\mathbf{P}(m, m+n-1) = \mathbf{P}^{n-1}$ by the induction hypothesis, and $\mathbf{P}(m+n-1, m+n) = \mathbf{P}$, whence $\mathbf{P}(m, m+n) = \mathbf{P}^n$. □

Notation

We will let $p_{ij}(n)$ denote the matrix entries of $\mathbf{P}^n$.

This allows us to express the probabilities at time n in terms of the initial probabilities.

This allows us to express the probabilities at time n in terms of the initial probabilities. Let $\pi_n(i) = \mathbb{P}(X_n = i)$ and consider the **probability vector** $\boldsymbol{\pi}_n$ whose ith entry is $\pi_n(i)$.

This allows us to express the probabilities at time n in terms of the initial probabilities. Let $\pi_n(i) = \mathbb{P}(X_n = i)$ and consider the **probability vector** $\boldsymbol{\pi}_n$ whose ith entry is $\pi_n(i)$.

Theorem

For every $n, m \geqslant 0$, $\boldsymbol{\pi}_{n+m} = \boldsymbol{\pi}_m \mathbf{P}^n$.

This allows us to express the probabilities at time n in terms of the initial probabilities. Let $\pi_n(i) = \mathbb{P}(X_n = i)$ and consider the **probability vector** $\boldsymbol{\pi}_n$ whose ith entry is $\pi_n(i)$.

Theorem

For every $n, m \geqslant 0$, $\boldsymbol{\pi}_{n+m} = \boldsymbol{\pi}_m \mathbf{P}^n$.

Proof.

By the partition rule,

$$\begin{aligned}\mathbb{P}(X_{m+n} = j) &= \sum_i \mathbb{P}(X_{m+n} = j \mid X_m = i)\mathbb{P}(X_m = i) \\ &= \sum_i p_{ij}(m, m+n)\pi_m(i)\end{aligned}$$

This allows us to express the probabilities at time n in terms of the initial probabilities. Let $\pi_n(i) = \mathbb{P}(X_n = i)$ and consider the **probability vector** $\boldsymbol{\pi}_n$ whose ith entry is $\pi_n(i)$.

Theorem

For every $n, m \geqslant 0$, $\boldsymbol{\pi}_{n+m} = \boldsymbol{\pi}_m \mathbf{P}^n$.

Proof.

By the partition rule,

$$\begin{aligned}\mathbb{P}(X_{m+n} = j) &= \sum_i \mathbb{P}(X_{m+n} = j \mid X_m = i)\mathbb{P}(X_m = i) \\ &= \sum_i p_{ij}(m, m+n)\pi_m(i)\end{aligned}$$

which in terms of matrices is the product

$$\boldsymbol{\pi}_{n+m} = \boldsymbol{\pi}_m \mathbf{P}(m, m+n) = \boldsymbol{\pi}_m \mathbf{P}^n$$

□

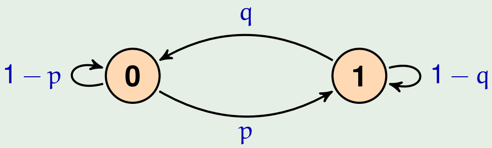
So in particular, $\pi_n = \pi_0 P^n$, so that the probabilities π_n at time n are the initial probabilities π_0 multiplied with the nth power of the transition matrix.

So in particular, $\pi_n = \pi_0 P^n$, so that the probabilities π_n at time n are the initial probabilities π_0 multiplied with the nth power of the transition matrix. The transition matrices carry most of the information in the Markov chain.

So in particular, $\pi_n = \pi_0 P^n$, so that the probabilities π_n at time n are the initial probabilities π_0 multiplied with the nth power of the transition matrix. The transition matrices carry most of the information in the Markov chain.

Example

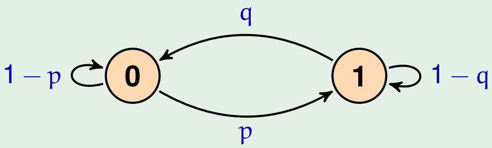
Consider the general 2-state Markov chain



So in particular, $\pi_n = \pi_0 \mathbf{P}^n$, so that the probabilities π_n at time n are the initial probabilities π_0 multiplied with the nth power of the transition matrix. The transition matrices carry most of the information in the Markov chain.

Example

Consider the general 2-state Markov chain



with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Example (Continued)

We proved earlier that

$$\pi_n(0) = (1 - p - q)^n \left(\pi_0(0) - \frac{q}{p+q}\right) + \frac{q}{p+q}$$
$$\pi_n(1) = (1 - p - q)^n \left(\pi_0(1) - \frac{p}{p+q}\right) + \frac{p}{p+q}$$

Example (Continued)

We proved earlier that

$$\pi_n(0) = (1-p-q)^n \left(\pi_0(0) - \frac{q}{p+q}\right) + \frac{q}{p+q}$$
$$\pi_n(1) = (1-p-q)^n \left(\pi_0(1) - \frac{p}{p+q}\right) + \frac{p}{p+q}$$

and we can use this to calculate the n-step transition matrix $\mathbf{P}^n$.

Example (Continued)

We proved earlier that

$$\pi_n(0) = (1-p-q)^n \left(\pi_0(0) - \frac{q}{p+q}\right) + \frac{q}{p+q}$$
$$\pi_n(1) = (1-p-q)^n \left(\pi_0(1) - \frac{p}{p+q}\right) + \frac{p}{p+q}$$

and we can use this to calculate the n-step transition matrix $\mathbf{P}^n$. Notice that for any 2×2 matrix $\boldsymbol{A}$:

$$(1,0)\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = (a_{00}, a_{01}) \qquad (0,1)\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = (a_{10}, a_{11})$$

Example (Continued)

We proved earlier that

$$\pi_n(0) = (1-p-q)^n \left(\pi_0(0) - \frac{q}{p+q}\right) + \frac{q}{p+q}$$
$$\pi_n(1) = (1-p-q)^n \left(\pi_0(1) - \frac{p}{p+q}\right) + \frac{p}{p+q}$$

and we can use this to calculate the n-step transition matrix $\mathbf{P}^n$. Notice that for any 2×2 matrix $\mathbf{A}$:

$$(1,0)\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = (a_{00}, a_{01}) \qquad (0,1)\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} = (a_{10}, a_{11})$$

whence setting $\pi_0(0) = 1$ and $\pi_0(0) = 0$ in turn we read off

$$\mathbf{P}^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$$

Stationary probability distributions

In the previous example, notice that if $2 > p + q > 0$, then $|1 - p - q| < 1$ and hence $(1 - p - q)^n \to 0$ as $n \to \infty$.

Stationary probability distributions

In the previous example, notice that if $2 > p + q > 0$, then $|1 - p - q| < 1$ and hence $(1 - p - q)^n \to 0$ as $n \to \infty$. Therefore as $n \to \infty$,

$$\mathbf{P}^n \to \mathbf{P}^\infty = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

Stationary probability distributions

In the previous example, notice that if $2 > p + q > 0$, then $|1 - p - q| < 1$ and hence $(1 - p - q)^n \to 0$ as $n \to \infty$. Therefore as $n \to \infty$,

$$\mathbf{P}^n \to \mathbf{P}^\infty = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

This matrix $\mathbf{P}^\infty$ has the property that for any choice of initial probabilities $\boldsymbol{\pi}_0 = (\pi_0(0), \pi_0(1))$,

$$\boldsymbol{\pi}_0 \mathbf{P}^\infty = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$

Stationary probability distributions

In the previous example, notice that if $2 > p + q > 0$, then $|1 - p - q| < 1$ and hence $(1 - p - q)^n \to 0$ as $n \to \infty$. Therefore as $n \to \infty$,

$$\mathbf{P}^n \to \mathbf{P}^\infty = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

This matrix $\mathbf{P}^\infty$ has the property that for any choice of initial probabilities $\boldsymbol{\pi}_0 = (\pi_0(0), \pi_0(1))$,

$$\boldsymbol{\pi}_0 \mathbf{P}^\infty = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$

The probability vector $\boldsymbol{\pi} = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ is **stationary**: $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$.

Stationary probability distributions

In the previous example, notice that if $2 > p + q > 0$, then $|1 - p - q| < 1$ and hence $(1 - p - q)^n \to 0$ as $n \to \infty$. Therefore as $n \to \infty$,

$$\mathbf{P}^n \to \mathbf{P}^\infty = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

This matrix $\mathbf{P}^\infty$ has the property that for any choice of initial probabilities $\boldsymbol{\pi}_0 = (\pi_0(0), \pi_0(1))$,

$$\boldsymbol{\pi}_0 \mathbf{P}^\infty = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$

The probability vector $\boldsymbol{\pi} = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ is **stationary**: $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. Indeed,

$$\left(\frac{q}{p+q}, \frac{p}{p+q}\right)\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$

Definition

Let $\mathbf{P}$ be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi\mathbf{P} = \pi$.

Definition

Let P be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi P = \pi$.

Questions

1. Do all (finite-state) Markov chains have steady state distributions?

Definition

Let $\mathbf{P}$ be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi\mathbf{P} = \pi$.

Questions

1. Do all (finite-state) Markov chains have steady state distributions?
2. If so, is there a unique steady state distribution?

Definition

Let $\mathbf{P}$ be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi\mathbf{P} = \pi$.

Questions

1. Do all (finite-state) Markov chains have steady state distributions?
2. If so, is there a unique steady state distribution?
3. If so, will any initial distribution converge to the steady state distribution?

Definition

Let P be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi P = \pi$.

Questions

1. Do all (finite-state) Markov chains have steady state distributions?
2. If so, is there a unique steady state distribution?
3. If so, will any initial distribution converge to the steady state distribution?

Answers

1. **Yes**! (but we will not prove it in this course)

Definition

Let P be the transition matrix of a finite-state Markov chain. A probability vector π is a **steady state distribution** if $\pi P = \pi$.

Questions

1. Do all (finite-state) Markov chains have steady state distributions?
2. If so, is there a unique steady state distribution?
3. If so, will any initial distribution converge to the steady state distribution?

Answers

1. **Yes**! (but we will not prove it in this course)
2. Not necessarily.

Definition

Let $\mathbf{P}$ be the transition matrix of a finite-state Markov chain. A probability vector $\boldsymbol{\pi}$ is a **steady state distribution** if $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$.

Questions

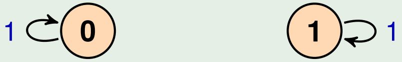
1. Do all (finite-state) Markov chains have steady state distributions?
2. If so, is there a unique steady state distribution?
3. If so, will any initial distribution converge to the steady state distribution?

Answers

1. **Yes**! (but we will not prove it in this course)
2. Not necessarily.
3. Not necessarily.

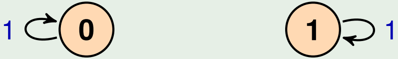
Example

Consider the following 2-state Markov chain



Example

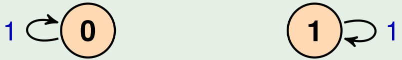
Consider the following 2-state Markov chain



$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Example

Consider the following 2-state Markov chain

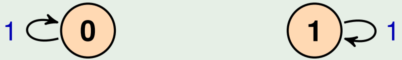


$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then clearly every π obeys $\pi = \pi\mathbf{P}$.

Example

Consider the following 2-state Markov chain



$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

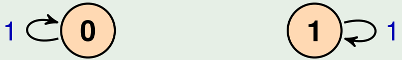
Then clearly every π obeys $\pi = \pi \mathbf{P}$.

Post-mortem

The problem here is that the Markov chain decomposes: not every state is "accessible" from every other state.

Example

Consider the following 2-state Markov chain



$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then clearly every π obeys $\pi = \pi P$.

Post-mortem

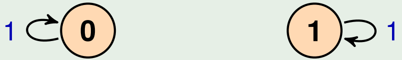
The problem here is that the Markov chain decomposes: not every state is "accessible" from every other state.

Definition

A state j is **accessible** from a state i, if for some $n \geqslant 0$, $p_{ij}(n) > 0$.

Example

Consider the following 2-state Markov chain



$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then clearly every π obeys $\pi = \pi P$.

Post-mortem

The problem here is that the Markov chain decomposes: not every state is "accessible" from every other state.

Definition

A state j is **accessible** from a state i, if for some $n \geqslant 0$, $p_{ij}(n) > 0$. A Markov chain is **irreducible** if any state is accessible from any other state; i.e., given any two states i, j, there is some $n \geqslant 0$ with $p_{ij}(n) > 0$.

Uniqueness of steady state distribution

Theorem

An irreducible finite-state Markov chain has a unique steady state distribution.

Uniqueness of steady state distribution

Theorem

An irreducible finite-state Markov chain has a unique steady state distribution.

Warning

If the Markov chain has an infinite (but still countable) number of states, then this is not true; although there are theorems guaranteeing the uniqueness of a steady state distribution in those cases as well.

Uniqueness of steady state distribution

Theorem

An irreducible finite-state Markov chain has a unique steady state distribution.

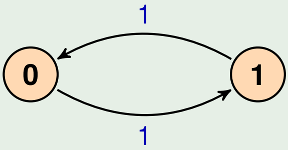
Warning

If the Markov chain has an infinite (but still countable) number of states, then this is not true; although there are theorems guaranteeing the uniqueness of a steady state distribution in those cases as well.

This still leaves the question of whether in a Markov chain with a unique steady state distribution, any initial distribution eventually tends to it.

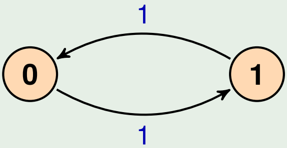
Example

Consider the following 2-state Markov chain



Example

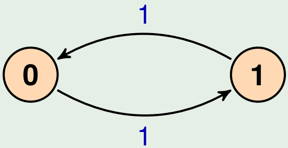
Consider the following 2-state Markov chain



$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Example

Consider the following 2-state Markov chain

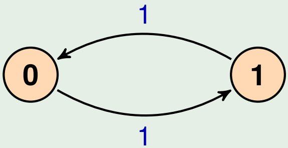


$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Then there is a unique steady state distribution $\boldsymbol{\pi} = \left(\frac{1}{2}, \frac{1}{2}\right)$, but no other distribution converges to it.

Example

Consider the following 2-state Markov chain



$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Then there is a unique steady state distribution $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$, but no other distribution converges to it.

Post-mortem

The problem here is that $\mathbf{P}^2$ is the identity matrix, so every distribution (except the steady state distribution) has "period" 2.

Periods

Definition

A state i is said to be **periodic** with period k if any return visit to i occurs in multiples of k time steps.

Periods

Definition

A state i is said to be **periodic** with period k if any return visit to i occurs in multiples of k time steps. More precisely, let

$$k_i = \gcd\{n \mid \mathbb{P}(X_n = i \mid X_0 = i) > 0\}$$

Periods

Definition

A state i is said to be **periodic** with period k if any return visit to i occurs in multiples of k time steps. More precisely, let

$$k_i = \gcd\{n \mid \mathbb{P}(X_n = i \mid X_0 = i) > 0\}$$

Then if $k_i > 1$, the state i is periodic with period k_i and if $k_i = 1$, the state i is **aperiodic**.

Periods

Definition

A state i is said to be **periodic** with period k if any return visit to i occurs in multiples of k time steps. More precisely, let

$$k_i = \text{gcd}\{n \mid \mathbb{P}(X_n = i \mid X_0 = i) > 0\}$$

Then if $k_i > 1$, the state i is periodic with period k_i and if $k_i = 1$, the state i is **aperiodic**. A Markov chain is said to be **aperiodic** if all states are aperiodic.

Periods

Definition

A state i is said to be **periodic** with period k if any return visit to i occurs in multiples of k time steps. More precisely, let

$$k_i = \gcd\{n \mid \mathbb{P}(X_n = i \mid X_0 = i) > 0\}$$

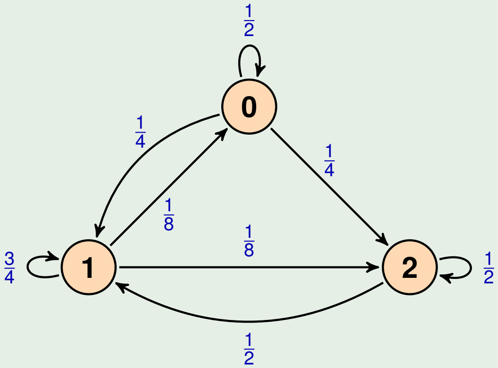
Then if $k_i > 1$, the state i is periodic with period k_i and if $k_i = 1$, the state i is **aperiodic**. A Markov chain is said to be **aperiodic** if all states are aperiodic.

Theorem

An irreducible, aperiodic, finite-state Markov chain has a unique steady state distribution π to which any initial distribution will eventually converge: for all π_0, $\pi_0 \mathbf{P}^n \to \pi$ as $n \to \infty$.

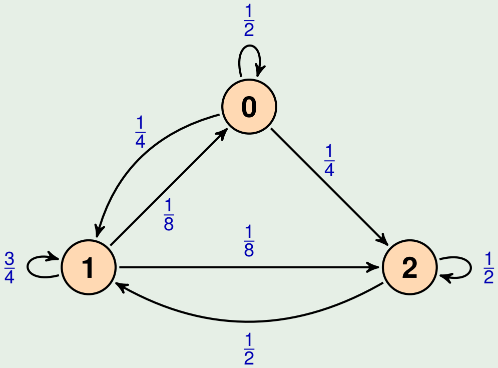
Example

Consider the following 3-state Markov chain



Example

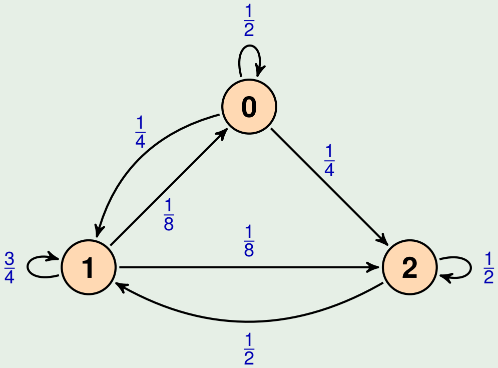
Consider the following 3-state Markov chain



$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Example

Consider the following 3-state Markov chain



$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{8} & \frac{3}{4} & \frac{1}{8} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Solving the equation $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ for $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$ with $\pi_0 + \pi_1 + \pi_2 = 1$, we find $\boldsymbol{\pi} = \left(\frac{2}{13}, \frac{8}{13}, \frac{3}{13}\right)$. Moreover, any initial distribution converges to it.

Example (Continued)

The reason is the limit $n \to \infty$ of $\mathbf{P}^n$ exists:

$$\mathbf{P}^n \to \begin{pmatrix} \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \end{pmatrix}$$

Example (Continued)

The reason is the limit $n \to \infty$ of $\mathbf{P}^n$ exists:

$$\mathbf{P}^n \to \begin{pmatrix} \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \end{pmatrix}$$

And hence for any (α, β, γ) with $\alpha + \beta + \gamma = 1$,

$$(\alpha, \beta, \gamma) \begin{pmatrix} \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \end{pmatrix} = (\alpha + \beta + \gamma)(\tfrac{2}{13}, \tfrac{8}{13}, \tfrac{3}{13}) = (\tfrac{2}{13}, \tfrac{8}{13}, \tfrac{3}{13})$$

Example (Continued)

The reason is the limit $n \to \infty$ of $\mathbf{P}^n$ exists:

$$\mathbf{P}^n \to \begin{pmatrix} \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \end{pmatrix}$$

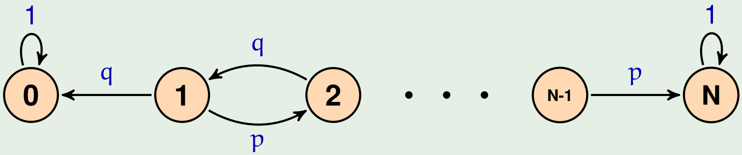
And hence for any (α, β, γ) with $\alpha + \beta + \gamma = 1$,

$$(\alpha, \beta, \gamma) \begin{pmatrix} \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \\ \frac{2}{13} & \frac{8}{13} & \frac{3}{13} \end{pmatrix} = (\alpha + \beta + \gamma)(\tfrac{2}{13}, \tfrac{8}{13}, \tfrac{3}{13}) = (\tfrac{2}{13}, \tfrac{8}{13}, \tfrac{3}{13})$$

It is actually enough to show that for some $n \geqslant 1$, $\mathbf{P}^n$ has no zero entries!

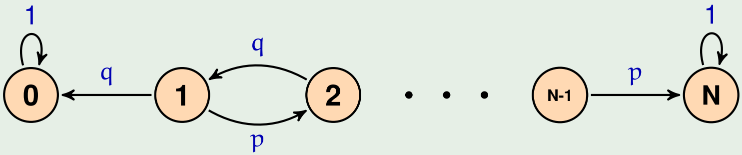
Example (Gambler's ruin – revisited)

Consider again the example of a random walk on $\{0, 1, \ldots, N\}$:



Example (Gambler's ruin – revisited)

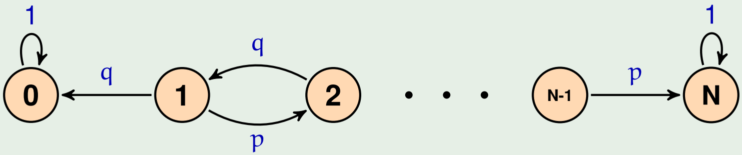
Consider again the example of a random walk on $\{0, 1, \ldots, N\}$:



States 0 and N are **absorbing**; i.e., $p_{00} = p_{NN} = 1$.

Example (Gambler's ruin – revisited)

Consider again the example of a random walk on $\{0, 1, \ldots, N\}$:



States 0 and N are **absorbing**; i.e., $p_{00} = p_{NN} = 1$.

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ q & 0 & p & 0 & \ldots & 0 \\ 0 & q & 0 & p & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & & q & 0 & p \\ 0 & 0 & \ldots & 0 & 0 & 1 \end{pmatrix}$$

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.
- $N = |\mathcal{S}| \gtrsim 8.42 \times 10^9$ as of Friday 16 March 2012.

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.
- $N = |\mathcal{S}| \gtrsim 8.42 \times 10^9$ as of Friday 16 March 2012.
- Let us write $i \to j$ if web page i has a link to web page j.

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.
- $N = |\mathcal{S}| \gtrsim 8.42 \times 10^9$ as of Friday 16 March 2012.
- Let us write $i \to j$ if web page i has a link to web page j.
- Set $b_i = |\{j \mid i \to j\}|$: the number of outlinks from i.

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.
- $N = |\mathcal{S}| \gtrsim 8.42 \times 10^9$ as of Friday 16 March 2012.
- Let us write $i \to j$ if web page i has a link to web page j.
- Set $b_i = |\{j \mid i \to j\}|$: the number of outlinks from i.
- The transition matrix $\mathbf{P}$ has entries (for $\delta \simeq 0.85$)

$$p_{ij} = \begin{cases} \frac{1-\delta}{N} + \frac{\delta}{b_i}, & i \to j \\ \frac{1-\delta}{N}, & i \not\to j \end{cases}$$

Example (Google's PageRank)

- Google's PageRank algorithm is a Markov chain!
- A random "surf" on the set $\mathcal{S}$ of all (public) web pages.
- $N = |\mathcal{S}| \gtrsim 8.42 \times 10^9$ as of Friday 16 March 2012.
- Let us write $i \to j$ if web page i has a link to web page j.
- Set $b_i = |\{j \mid i \to j\}|$: the number of outlinks from i.
- The transition matrix $\mathbf{P}$ has entries (for $\delta \simeq 0.85$)

$$p_{ij} = \begin{cases} \frac{1-\delta}{N} + \frac{\delta}{b_i}, & i \to j \\ \frac{1-\delta}{N}, & i \not\to j \end{cases}$$

$$\begin{aligned} \sum_j p_{ij} &= \sum_{j \leftarrow i} \left(\frac{\delta}{b_i} + \frac{1-\delta}{N} \right) + \sum_{j \not\leftarrow i} \frac{1-\delta}{N} \\ &= b_i \left(\frac{\delta}{b_i} + \frac{1-\delta}{N} \right) + (N - b_i) \frac{1-\delta}{N} \\ &= 1 \end{aligned}$$

Example (Google's PageRank – continued)

- Since $p_{ij} > 0$, the Markov chain is irreducible and aperiodic.

Example (Google's PageRank – continued)

- Since $p_{ij} > 0$, the Markov chain is irreducible and aperiodic.
- Therefore there is a unique steady state distribution to which every initial distribution converges to.

Example (Google's PageRank – continued)

- Since $p_{ij} > 0$, the Markov chain is irreducible and aperiodic.
- Therefore there is a unique steady state distribution to which every initial distribution converges to.
- This steady state distribution is the PageRank!

Example (Google's PageRank – continued)

- Since $p_{ij} > 0$, the Markov chain is irreducible and aperiodic.
- Therefore there is a unique steady state distribution to which every initial distribution converges to.
- This steady state distribution is the PageRank!
- The PageRank $\boldsymbol{\pi} = (\pi_j)$ obeys the equation

$$\pi_j = \frac{1-\delta}{N} + \delta \sum_{i \to j} \frac{\pi_i}{b_i}$$

Example (Google's PageRank – continued)

- Since $p_{ij} > 0$, the Markov chain is irreducible and aperiodic.
- Therefore there is a unique steady state distribution to which every initial distribution converges to.
- This steady state distribution is the PageRank!
- The PageRank $\boldsymbol{\pi} = (\pi_j)$ obeys the equation

$$\pi_j = \frac{1-\delta}{N} + \delta \sum_{i \to j} \frac{\pi_i}{b_i}$$

- It can be solved by iteration, which for large N converges relatively quickly.

Example

Alice, Bob and Sergei each have a webpage in their home network.

Example

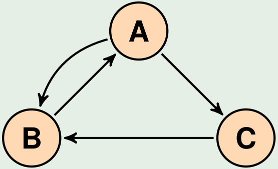
Alice, Bob and Sergei each have a webpage in their home network. Alice's page points to both Bob's and Sergei's, whereas Bob's page only points back to Alice's and Sergei's only points to Bob's.

Example

Alice, Bob and Sergei each have a webpage in their home network. Alice's page points to both Bob's and Sergei's, whereas Bob's page only points back to Alice's and Sergei's only points to Bob's. *What are their PageRanks?*

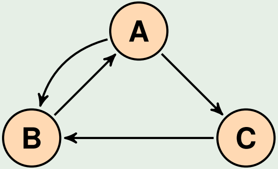
Example

Alice, Bob and Sergei each have a webpage in their home network. Alice's page points to both Bob's and Sergei's, whereas Bob's page only points back to Alice's and Sergei's only points to Bob's. *What are their PageRanks?*



Example

Alice, Bob and Sergei each have a webpage in their home network. Alice's page points to both Bob's and Sergei's, whereas Bob's page only points back to Alice's and Sergei's only points to Bob's. *What are their PageRanks?*

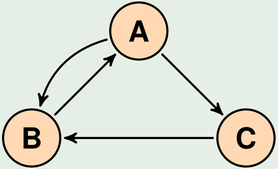


$$N = 3 \qquad b_A = 2 \qquad b_B = b_C = 1$$

$$\mathbf{P} = \begin{pmatrix} \frac{1-\delta}{3} & \frac{1-\delta}{3} + \frac{\delta}{2} & \frac{1-\delta}{3} + \frac{\delta}{2} \\ \frac{1-\delta}{3} + \delta & \frac{1-\delta}{3} & \frac{1-\delta}{3} \\ \frac{1-\delta}{3} & \frac{1-\delta}{3} + \delta & \frac{1-\delta}{3} \end{pmatrix}$$

Example

Alice, Bob and Sergei each have a webpage in their home network. Alice's page points to both Bob's and Sergei's, whereas Bob's page only points back to Alice's and Sergei's only points to Bob's. *What are their PageRanks?*



$$N = 3 \qquad b_A = 2 \qquad b_B = b_C = 1$$

$$\mathbf{P} = \begin{pmatrix} \frac{1-\delta}{3} & \frac{1-\delta}{3} + \frac{\delta}{2} & \frac{1-\delta}{3} + \frac{\delta}{2} \\ \frac{1-\delta}{3} + \delta & \frac{1-\delta}{3} & \frac{1-\delta}{3} \\ \frac{1-\delta}{3} & \frac{1-\delta}{3} + \delta & \frac{1-\delta}{3} \end{pmatrix}$$

$$(\delta = 0.85) \qquad \mathbf{P} = \begin{pmatrix} 0.05 & 0.475 & 0.475 \\ 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \end{pmatrix} \implies \boldsymbol{\pi}^T \simeq \begin{pmatrix} 0.39 \\ 0.40 \\ 0.21 \end{pmatrix}$$

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** P and the n-step transition matrix is P^n

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** $\mathbf{P}$ and the n-step transition matrix is $\mathbf{P}^n$
- The probability distribution $\boldsymbol{\pi}_m$ at time m obeys $\boldsymbol{\pi}_{m+n} = \boldsymbol{\pi}_m \mathbf{P}^n$ for all $m, n \geqslant 0$

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** $\mathbf{P}$ and the n-step transition matrix is $\mathbf{P}^n$
- The probability distribution $\boldsymbol{\pi}_m$ at time m obeys $\boldsymbol{\pi}_{m+n} = \boldsymbol{\pi}_m \mathbf{P}^n$ for all $m, n \geqslant 0$
- A probability distribution $\boldsymbol{\pi}$ is a **steady state distribution** if $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** P and the n-step transition matrix is P^n
- The probability distribution π_m at time m obeys $\pi_{m+n} = \pi_m P^n$ for all $m, n \geqslant 0$
- A probability distribution π is a **steady state distribution** if $\pi P = \pi$
- Finite-state Markov chains always have steady state distributions.

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** P and the n-step transition matrix is P^n
- The probability distribution π_m at time m obeys $\pi_{m+n} = \pi_m P^n$ for all $m, n \geqslant 0$
- A probability distribution π is a **steady state distribution** if $\pi P = \pi$
- Finite-state Markov chains always have steady state distributions.
- A necessary and sufficient condition for a finite-state Markov chain to have a unique steady state distribution to which all distributions converge is that for some n, P^n has no zero entries.

Summary

- Temporally homogeneous Markov chains are characterised by an stochastic **transition matrix** P and the n-step transition matrix is P^n
- The probability distribution π_m at time m obeys $\pi_{m+n} = \pi_m P^n$ for all $m, n \geqslant 0$
- A probability distribution π is a **steady state distribution** if $\pi P = \pi$
- Finite-state Markov chains always have steady state distributions.
- A necessary and sufficient condition for a finite-state Markov chain to have a unique steady state distribution to which all distributions converge is that for some n, P^n has no zero entries.
- Google's PageRank algorithm is a Markov chain!