Introduction to Markov Chains

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

Introduction

Markov Processes and Markov Chains

- We study random processes that retain no memory of where they have been in the past.
- This means that only the current state of the process can influence where it goes next.
- Such a process is called a **Markov process**.
- We deal exclusively with the case where the process can assume only a finite or countable set of states, when it is referred to as a **Markov chain**.

Discrete and Continuous Time

- We consider chains both in **discrete time**

$$n \in \mathbb{Z}^+ = \{0, 1, 2, \dots\}$$

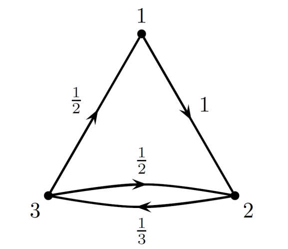
  and **continuous time**

$$t \in \mathbb{R}^+ = [0, \infty).$$

- The letters n, m, k will always denote integers.
- The letters t and s will refer to real numbers.
- Thus, we write:
  - $(X_n)_{n \geq 0}$ for a discrete-time process;
  - $(X_t)_{t \geq 0}$ for a continuous-time process.

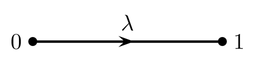
Example: Discrete Time

- We move from state 1 to state 2 with probability 1.
- From state 3, we move either to 1 or to 2 with equal probability 1/2.
- From 2, we jump to 3 with probability 1/3, otherwise stay at 2.



- We might have drawn a loop from 2 to itself with label 2/3.
- Since the total probability on jumping from 2 must equal 1, this does not convey any more information.
- So one may leave loops out.

Example: Continuous Time

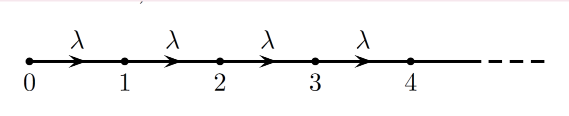


- When in state 0, we wait for a random time with exponential distribution of parameter $\lambda \in (0, \infty)$, then jump to 1.
- Thus the density function of the waiting time T is given by

$$f_T(t) = \lambda e^{-\lambda t}, \quad \text{for } t \geq 0.$$

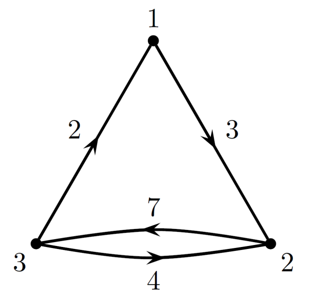
- We write $T \sim E(\lambda)$ for short.

Example: Poisson Process of Rate λ



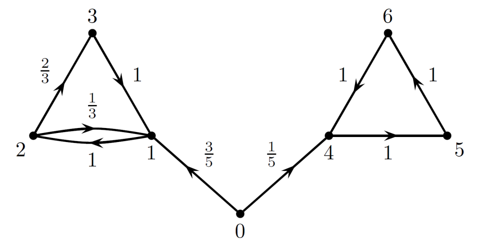
- Here, when we get to 1, we do not stop but, after another independent exponential time of parameter λ, jump to 2, and so on.
- The resulting process is called the **Poisson process of rate** λ.

Example: Continuous Time



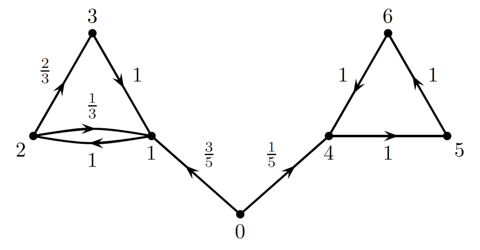
- In state 3, we take two independent exponential times $T_1 \sim E(2)$ and $T_2 \sim E(4)$.
  - If T_1 is the smaller, we go to 1 after time T_1;
  - If T_2 is the smaller, we go to 2 after time T_2.
- The rules for states 1 and 2 are as given in the preceding examples.
- We will show later that:
  - The time spent in 3 is exponential of parameter $2 + 4 = 6$;
  - The probability of jumping from 3 to 1 is $\frac{2}{2+4} = \frac{1}{3}$.

Example: Discrete Time



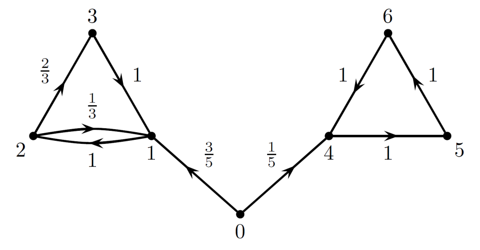
- The states may be partitioned into **communicating classes**, namely $\{0\}$, $\{1, 2, 3\}$ and $\{4, 5, 6\}$.
- Two of these classes are **closed**, meaning that you cannot escape.
- The closed classes here are **recurrent**, meaning that you return again and again to every state.
- The class $\{0\}$ is **transient**.
- The class $\{4, 5, 6\}$ is **periodic**, but $\{1, 2, 3\}$ is not.

Example: Discrete Time (Cont'd)



- The following hold:
  (a) Starting from 0, the probability of hitting 6 is $\frac{1}{4}$.
  (b) Starting from 1, the probability of hitting 3 is 1.
  (c) Starting from 1, it takes on average three steps to hit 3.
  (d) Starting from 1, the long-run proportion of time spent in 2 is $\frac{3}{8}$.

Example: Discrete Time (Cont'd)



- Let $p_{ij}^{(n)}$ be the probability of being in state j after n steps, when starting from state i.
- Then we also have:
  (e) $\lim_{n\to\infty} p_{01}^{(n)} = \frac{9}{32}$;
  (f) $p_{04}^{(n)}$ does not converge as $n \to \infty$;
  (g) $\lim_{n\to\infty} p_{04}^{(3n)} = \frac{1}{124}$.

Subsection 2

Definition and Basic Properties

State Spaces and Distributions

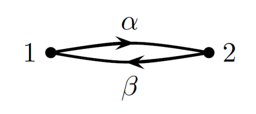
- Let I be a countable set.
- Each $i \in I$ is called a **state** and I is called the **state space**.
- We say that $\lambda = (\lambda_i : i \in I)$ is a **measure** on I if $0 \leq \lambda_i < \infty$, for all $i \in I$.
- If, in addition the **total mass** $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a **distribution**.
- We work throughout with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- A **random variable** X with values in I is a function $X : \Omega \to I$.
- Suppose we set $\lambda_i = \mathbb{P}(X = i) = \mathbb{P}(\{\omega : X(\omega) = i\})$.
- Then λ defines a distribution, the **distribution of** X.
- We think of X as modelling a random state which takes the value i with probability λ_i.

Stochastic Matrices

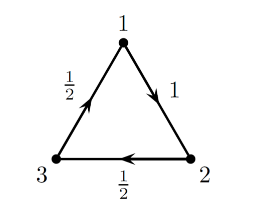
- We say that a matrix $P = (p_{ij} : i, j \in I)$ is **stochastic** if every row $(p_{ij} : j \in I)$ is a distribution.
- There is a one-to-one correspondence between stochastic matrices P and the sort of diagrams described in the Introduction.

  Example:

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix};$$



$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$



Markov Chains

- We say that $(X_n)_{n\geq 0}$ is a **Markov chain** with **initial distribution** λ and **transition matrix** P if:
  (i) X_0 has distribution λ;
  (ii) For n$\geq$ 0, conditional on $X_n = i$, X_{n+1} has distribution $(p_{ij} : j \in I)$ and is independent of $X_0, \ldots, X_{n-1}$.
- More explicitly, these conditions state that, for $n \geq 0$ and $i_1, \ldots, i_{n+1} \in I$,
  (i) $\mathbb{P}(X_0 = i_1) = \lambda_{i_1}$;
  (ii) $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_1, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$.
- We say that $(X_n)_{n\geq 0}$ is Markov(λ, P) for short.
- If $(X_n)_{0\leq n\leq N}$ is a finite sequence of random variables satisfying Conditions (i) and (ii), for $n = 0, \ldots, N-1$, then we again say $(X_n)_{0\leq n\leq N}$ is Markov(λ, P).

Characterization Theorem

Theorem

A discrete-time random process $(X_n)_{0\leq n\leq N}$ is Markov(λ, P) if and only if for all $i_0, i_1, \ldots, i_N \in I$,

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N}.$$

- Suppose $(X_n)_{0\leq n\leq N}$ is Markov(λ, P). Then

$$\begin{aligned}&\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N)\\ &= \mathbb{P}(X_0 = i_0)\mathbb{P}(X_1 = i_1 | X_0 = i_0)\\ &\qquad \cdots \mathbb{P}(X_N = i_N | X_0 = i_0, \ldots, X_{N-1} = i_{N-1})\\ &= \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{N-1} i_N}.\end{aligned}$$

Characterization Theorem (Converse)

- On the other hand, suppose the equation holds for N.
  By summing both sides over $i_N \in I$ and using $\sum_{j \in I} p_{ij} = 1$, we see that the equation holds for $N-1$.
  By induction, for all $n = 0, 1, \ldots, N$,
  $$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$
  In particular:
  - $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$;
  - For $n = 0, 1, \ldots, N-1$,
  $$\begin{aligned} \mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) &= \frac{\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1})}{\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n)} \\ &= p_{i_n i_{n+1}}. \end{aligned}$$

  So $(X_n)_{0 \le n \le N}$ is Markov(λ, P).

Markov Property

- Write $\delta_i = (\delta_{ij} : j \in I)$ for the **unit mass** at i, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem (Markov Property)

Let $(X_n)_{n \geq 0}$ be Markov(λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n \geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$.

- We have to show that, for any event A determined by $X_0, \ldots, X_m$,

$$\begin{aligned} & \mathbb{P}(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A | X_m = i) \\ & = \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} \mathbb{P}(A | X_m = i). \end{aligned}$$

Then the result follows by the preceding theorem.

Markov Property (Cont'd)

- First consider the case of elementary events

$$A = \{X_0 = i_0, \ldots, X_m = i_m\}.$$

In that case we have to show

$$\begin{aligned} &\frac{\mathbb{P}(X_0=i_0,\ldots,X_{m+n}=i_{m+n} \text{ and } i=i_m)}{\mathbb{P}(X_m=i)} \\ &= \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} \times \frac{\mathbb{P}(X_0=i_0,\ldots,X_m=i_m \text{ and } i=i_m)}{\mathbb{P}(X_m=i)}. \end{aligned}$$

This is true by the preceding theorem.

In general, any event A determined by $X_0, \ldots, X_m$ may be written as a countable disjoint union of elementary events $A = \bigcup_{k=1}^{\infty} A_k$.

In this case, the desired identity for A follows by summing up the corresponding identities for A_k.

Matrix Notation

- We regard P as a matrix whose entries are indexed by $I \times I$.
- We regard distributions and measures λ as row vectors whose components are indexed by I.
- When I is finite we will often label the states $1, 2, \ldots, N$.
- In this case, λ will be an N-vector and P an $N \times N$-matrix.
- For finite objects, matrix multiplication is a familiar operation.

$$(\lambda P)_j = \sum_{i=1}^{N} \lambda_i p_{ij}, \quad (P^2)_{ik} = \sum_{j=1}^{N} p_{ij} p_{jk}.$$

Matrix Notation (Cont'd)

- We extend matrix multiplication to the general case.
- We define a new measure λP and a new matrix P^2 by

$$(\lambda P)_j = \sum_{i \in I} \lambda_i p_{ij}, \quad (P^2)_{ik} = \sum_{j \in I} p_{ij} p_{jk}.$$

- We define P^n similarly for any n.
- We agree that P^0 is the identity matrix I, where

$$(I)_{ij} = \delta_{ij}.$$

- We write $p_{ij}^{(n)} = (P^n)_{ij}$, for the (i, j) entry in P^n.

Conditional Probability $\mathbb{P}_i$

- In the case where $\lambda_i > 0$ we shall write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A|X_0 = i)$.
- By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n\geq 0}$ is Markov(δ_i, P).
- So the behavior of $(X_n)_{n\geq 0}$ under $\mathbb{P}_i$ does not depend on λ.

Transition Probabilities

Theorem

Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then, for all $n, m \geq 0$,

(i) $\mathbb{P}(X_n = j) = (\lambda P^n)_j$;

(ii) $\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j | X_m = i) = p_{ij}^{(n)}$.

(i) By a previous theorem,

$$\begin{aligned} \mathbb{P}(X_n = j) &= \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0 \in I} \cdots \sum_{i_{n-1} \in I} \lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} \\ &= (\lambda P^n)_j. \end{aligned}$$

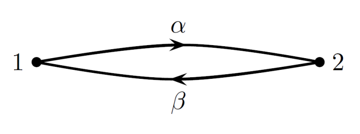
(ii) By the Markov property, conditional on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P). So we just take $\lambda = \delta_i$ in Part (i).

- We call $p_{ij}^{(n)}$ the ***n*-step transition probability from *i* to *j***.

Example

- The most general two-state chain has transition matrix of the form

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$



- We exploit the relation $P^{n+1} = P^n P$ to write

$$p_{11}^{(n+1)} = p_{12}^{(n)}\beta + p_{11}^{(n)}(1-\alpha).$$

- We also know that

$$p_{11}^{(n)} + p_{12}^{(n)} = \mathbb{P}_1(X_n = 1 \text{ or } 2) = 1.$$

Example (Cont'd)

- We wrote
$$p_{11}^{(n+1)} = p_{12}^{(n)}\beta + p_{11}^{(n)}(1-\alpha),$$
$$p_{11}^{(n)} + p_{12}^{(n)} = 1.$$
- By eliminating $p_{12}^{(n)}$ we get a recurrence relation for $p_{11}^{(n)}$,
$$p_{11}^{(n+1)} = (1-\alpha-\beta)p_{11}^{(n)} + \beta, \quad p_{11}^{(0)} = 1.$$
- This has a unique solution
$$p_{11}^{(n)} = \begin{cases} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n, & \text{for } \alpha+\beta>0 \\ 1, & \text{for } \alpha+\beta=0. \end{cases}$$

Example: Virus Mutation

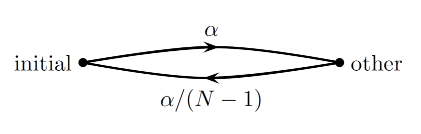
- Suppose a virus can exist in N different strains.
- In each generation it either stays the same, or with probability α mutates to another strain, which is chosen at random.
- We compute the probability that the strain in the n-th generation is the same as that in the 0-th generation.
- We could model this process as an N-state chain.
- The $N \times N$ transition matrix P given by

$$p_{ii} = 1 - \alpha, \quad p_{ij} = \frac{\alpha}{N-1}, \text{ for } i \neq j.$$

- Then the probability we seek is found by computing $p_{11}^{(n)}$.
- In this example there is a much simpler approach, which relies on exploiting the symmetry present in the mutation rules.

Example: Virus Mutation (Cont'd)

- At any time a transition is made:
  - From the initial state to another with probability α;
  - From another state to the initial state with probability $\frac{\alpha}{N-1}$.
- Thus, we have a two-state chain with the depicted diagram.



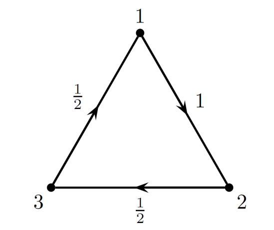
- By putting $\beta = \frac{\alpha}{N-1}$ in the preceding example, we find

$$\begin{aligned} p_{11}^{(n)} &= \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n \\ &= \frac{\frac{\alpha}{N-1}}{\alpha+\frac{\alpha}{N-1}} + \frac{\alpha}{\alpha+\frac{\alpha}{N-1}}\left(1-\alpha-\frac{\alpha}{N-1}\right)^n \\ &= \frac{1}{N} + \left(1-\frac{1}{N}\right)\left(1-\frac{\alpha N}{N-1}\right)^n. \end{aligned}$$

Example

- Consider the three-state chain shown.
- It has transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$



- We want to find a general formula for $p_{11}^{(n)}$.
- First we compute the eigenvalues of P.
- Its characteristic equation is

$$\det(x - P) = 0$$
$$x(x - \tfrac{1}{2})^2 - \tfrac{1}{4} = 0$$
$$\tfrac{1}{4}(x - 1)(4x^2 + 1) = 0.$$

- So the eigenvalues are 1, $\frac{i}{2}$ and $-\frac{i}{2}$.

Example (Cont'd)

- It follows that P is diagonalizable with

$$P = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{i}{2} & 0 \\ 0 & 0 & -\frac{i}{2} \end{pmatrix} U^{-1},$$

  for some invertible matrix U.
- So we get $P^n = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & (\frac{i}{2})^n & 0 \\ 0 & 0 & (-\frac{i}{2})^n \end{pmatrix} U^{-1}$.
- We conclude that $p_{11}^{(n)}$ has the form

$$p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n,$$

  for some constants a, b and c.

Example (Cont'd)

- We found that $p_{11}^{(n)}$ has the form

$$p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n,$$

for some constants a, b and c.

- The answer we want is real and

$$\left(\pm\frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n e^{\pm in\pi/2} = \left(\frac{1}{2}\right)^n \left(\cos\frac{n\pi}{2} \pm i\sin\frac{n\pi}{2}\right).$$

- So it makes sense to rewrite $p_{11}^{(n)}$ in the form

$$p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left\{\beta\cos\frac{n\pi}{2} + \gamma\sin\frac{n\pi}{2}\right\}$$

for constants α, β and γ.

Example (Conclusion)

- The first few values of $p_{11}^{(n)}$ are easy to write down.
- So we get equations to solve for α, β and γ:

$$\begin{array}{rcl} 1 & = & p_{11}^{(0)} = \alpha + \beta; \\ 0 & = & p_{11}^{(1)} = \alpha + \frac{1}{2}\gamma; \\ 0 & = & p_{11}^{(2)} = \alpha - \frac{1}{4}\beta. \end{array}$$

- So we get $\alpha = \frac{1}{5}$, $\beta = \frac{4}{5}$, $\gamma = -\frac{2}{5}$.
- It follows that

$$p_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left\{\frac{4}{5}\cos\frac{n\pi}{2} - \frac{2}{5}\sin\frac{n\pi}{2}\right\}.$$

The General Method

- The following method may in principle be used to find a formula for $p_{ij}^{(n)}$ for any M-state chain and any states i and j.
  (i) Compute the eigenvalues $\lambda_1, \ldots, \lambda_M$ of P by solving the characteristic equation.
  (ii) If the eigenvalues are distinct, then $p_{ij}^{(n)}$ has the form
  $$p_{ij}^{(n)} = a_1\lambda_1^n + \cdots + a_M\lambda_M^n,$$
  for some constants $a_1, \ldots, a_M$ (depending on i and j).
  If an eigenvalue λ is repeated (once, say) then the general form includes the term $(an + b)\lambda^n$.
  (iii) As roots of a polynomial with real coefficients, complex eigenvalues will come in conjugate pairs and these are best written using sine and cosine, as in the preceding example.

Subsection 3

Class Structure

Communicating Classes of a Chain

- We say that i **leads to** j, written $i \to j$, if
$$\mathbb{P}_i(X_n = j \text{ for some } n \geq 0) > 0.$$
- We say i **communicates with** j, written $i \leftrightarrow j$, if
$$i \to j \quad \text{and} \quad j \to i.$$

A Characterization Theorem

Theorem

For distinct states i and j the following are equivalent:

(i) $i \to j$;

(ii) $p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} > 0$, for some $i_1, i_2, \ldots, i_n$, with $i_1 = i$ and $i_n = j$;

(iii) $p_{ij}^{(n)} > 0$, for some $n \geq 0$.

- Observe that

$$p_{ij}^{(n)} \leq \mathbb{P}_i(X_n = j \text{ for some } n \geq 0) \leq \sum_{n=0}^{\infty} p_{ij}^{(n)}.$$

This proves the equivalence of (i) and (iii).

We also have $p_{ij}^{(n)} = \sum_{i_2,\ldots,i_{n-1}} p_{ii_2} p_{i_2 i_3} \cdots p_{i_{n-1} j}$.

So (ii) and (iii) are equivalent.

Closed, Absorbing and Irreducible Classes

- It is clear from (ii) that $i \to j$ and $j \to k$ imply $i \to k$.

  Also $i \to i$ for any state i.

  So $\leftrightarrow$ satisfies the conditions for an equivalence relation on I.

  Thus $\leftrightarrow$ partitions I into **communicating classes**.
- We say that a class C is **closed** if

  $$i \in C \quad \text{and} \quad i \to j \quad \text{imply} \quad j \in C.$$

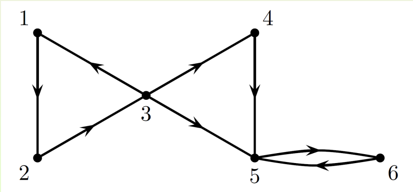
  Thus, a closed class is one from which there is no escape.
- A state i is **absorbing** if $\{i\}$ is a closed class.
- A chain or transition matrix P, where the set I of states is a single class, is called **irreducible**.

Example

- Find the communicating classes associated to the stochastic matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

- The solution is obvious from the diagram.
- The classes are $\{1,2,3\}$, $\{4\}$ and $\{5,6\}$.
- Only $\{5,6\}$ is closed.



Subsection 4

Hitting Times and Absorption Probabilities

Hitting Times

- Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P.
- The **hitting time** of a subset A of I is the random variable $H^A : \Omega \to \{0,1,2,\ldots\} \cup \{\infty\}$ given by

$$H^A(\omega) = \inf \{n \geq 0 : X_n(\omega) \in A\},$$

  where we agree that the infimum of the empty set $\emptyset$ is ∞.
- The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is then

$$h_i^A = \mathbb{P}_i(H^A < \infty).$$

Absorption Probabilities

- When A is a closed class,

$$h_i^A = \mathbb{P}_i(H^A < \infty)$$

is called the **absorption probability**.

- The mean time taken for $(X_n)_{n \geq 0}$ to reach A is given by

$$k_i^A = \mathbb{E}_i(H^A) = \sum_{n<\infty} n\mathbb{P}_i(H^A = n) + \infty\mathbb{P}_i(H^A = \infty).$$

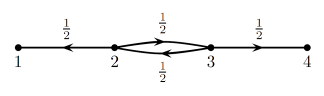
- We shall often write less formally

$$h_i^A = \mathbb{P}_i(\text{hit } A), \quad k_i^A = \mathbb{E}_i(\text{time to hit } A).$$

- These quantities can be calculated explicitly by means of certain linear equations associated with the transition matrix P.

Example

- Consider the chain with the following diagram:



Starting from 2, we calculate the probability of absorption in 4.

We also calculate the time until the chain is absorbed in 1 or 4.

Introduce $h_i = \mathbb{P}_i(\text{hit } 4)$, $k_i = \mathbb{E}_i(\text{time to hit } \{1,4\})$.

Clearly, $h_1 = 0$, $h_4 = 1$ and $k_1 = k_4 = 0$.

Suppose now that we start at 2.

Consider the situation after making one step.
  - With probability $\frac{1}{2}$ we jump to 1;
  - With probability $\frac{1}{2}$ we jump to 3.

So

$$\begin{array}{rcl} h_2 & = & \frac{1}{2}h_1 + \frac{1}{2}h_3, \\ k_2 & = & 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3. \end{array}$$

Example (Cont'd)

- We got

$$\begin{aligned} h_2 &= \tfrac{1}{2}h_1 + \tfrac{1}{2}h_3, \\ k_2 &= 1 + \tfrac{1}{2}k_1 + \tfrac{1}{2}k_3. \end{aligned}$$

Similarly,

$$\begin{aligned} h_3 &= \tfrac{1}{2}h_2 + \tfrac{1}{2}h_4, \\ k_3 &= 1 + \tfrac{1}{2}k_2 + \tfrac{1}{2}k_4. \end{aligned}$$

Hence

$$\begin{aligned} h_2 &= \tfrac{1}{2}h_3 = \tfrac{1}{2}(\tfrac{1}{2}h_2 + \tfrac{1}{2}), \\ k_2 &= 1 + \tfrac{1}{2}k_3 = 1 + \tfrac{1}{2}(1 + \tfrac{1}{2}k_2). \end{aligned}$$

So, starting from 2:

The probability of hitting 4 is $\frac{1}{3}$;

The mean time to absorption is 2.

Hitting Probabilities

Theorem

The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations

$$\begin{cases} h_i^A = 1, & \text{for } i \in A, \\ h_i^A = \sum_{j \in I} p_{ij} h_j^A & \text{for } i \notin A. \end{cases}$$

Minimality means that if $x = (x_i : i \in I)$ is another solution with $x_i \geq 0$, for all i, then $x_i \geq h_i^A$, for all i.

- First we show that h^A satisfies the system.
  Suppose $X_0 = i \in A$. Then $H^A = 0$. So $h_i^A = 1$.
  Suppose $X_0 = i \notin A$. Then $H^A \geq 1$.

Hitting Probabilities (Cont'd)

- By the Markov property,

$$\mathbb{P}_i(H^A < \infty | X_1 = j) = \mathbb{P}_j(H^A < \infty) = h_j^A.$$

Moreover,

$$\begin{aligned} h_i^A &= \mathbb{P}_i(H^A < \infty) \\ &= \textstyle\sum_{j \in I} \mathbb{P}_i(H^A < \infty, X_1 = j) \\ &= \textstyle\sum_{j \in I} \mathbb{P}_i(H^A < \infty | X_1 = j) \mathbb{P}_i(X_1 = j) \\ &= \textstyle\sum_{j \in I} p_{ij} h_j^A. \end{aligned}$$

Suppose, now, that $x = (x_i : i \in I)$ is a solution of the system.

For $i \in A$, $h_i^A = x_i = 1$.

Hitting Probabilities (Cont'd)

- Suppose $i \notin A$. Then

$$x_i = \sum_{j \in I} p_{ij} x_j = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} x_j.$$

Substitute for x_j to obtain

$$\begin{aligned} x_i &= \textstyle\sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} (\sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k) \\ &= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \textstyle\sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} x_k. \end{aligned}$$

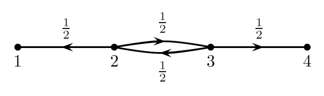
By repeated substitution for x in the final term we obtain after n steps

$$\begin{aligned} x_i = {} & \mathbb{P}_i(X_1 \in A) + \cdots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\ & + \textstyle\sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} x_{j_n}. \end{aligned}$$

Now if x is non-negative, so is the last term on the right.
Moreover, the remaining terms sum to $\mathbb{P}_i(H^A \le n)$.
So $x_i \ge \mathbb{P}_i(H^A \le n)$, for all n.
Then $x_i \ge \lim_{n \to \infty} \mathbb{P}_i(H^A \le n) = \mathbb{P}_i(H^A < \infty) = h_i$.

Example Revisited

- Consider again the chain shown.



The system of linear equations for $h = h^{\{4\}}$ are given by

$$\begin{array}{rcl} h_4 & = & 1, \\ h_2 & = & \frac{1}{2}h_1 + \frac{1}{2}h_3, \quad h_3 = \frac{1}{2}h_2 + \frac{1}{2}h_4. \end{array}$$

So

$$\begin{array}{rcl} h_2 & = & \frac{1}{2}h_1 + \frac{1}{2}(\frac{1}{2}h_2 + \frac{1}{2}), \\ h_2 & = & \frac{1}{3} + \frac{2}{3}h_1, \quad h_3 = \frac{2}{3} + \frac{1}{3}h_1. \end{array}$$

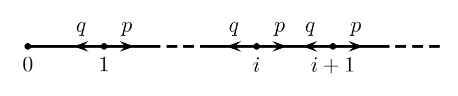
The value of h_1 is not determined by the system.

However, the minimality condition now makes us take $h_1 = 0$.

So we recover $h_2 = \frac{1}{3}$.

Example: Gambler's Ruin

- Consider the following Markov chain with $0 < p = 1 - q < 1$.



The transition probabilities are

$$p_{00} = 1, \quad p_{i,i-1} = q, \quad p_{i,i+1} = p, \quad \text{for } i = 1, 2, \ldots.$$

Imagine that we enter a casino with a fortune of \$i and gamble, \$1 at a time, with:

- Probability p of doubling our stake;
- Probability q of losing it.

The resources of the casino are regarded as infinite.

So there is no upper limit to our fortune.

We compute the probability that we go bust.

Example: Gambler's Ruin (Cont'd)

- Set $h_i = \mathbb{P}_i(\text{hit } 0)$.

  Then h is the minimal non-negative solution to

$$\begin{aligned} h_0 &= 1, \\ h_i &= ph_{i+1} + qh_{i-1}, \text{ for } i = 1, 2, \ldots. \end{aligned}$$

  Suppose $p \neq q$.

  Then the recurrence has a general solution

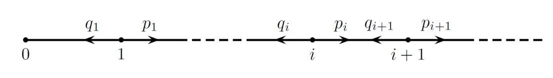
$$h_i = A + B \left(\frac{q}{p} \right)^i.$$

Example: Gamblers' Ruin (Cont'd)

- For $p \neq q$, we have $h_i = A + B(\frac{q}{p})^i$.
  - Suppose $p < q$.
    Since $0 \leq h_i \leq 1$, $B = 0$. So $h_i = 1$, for all i.
  - Suppose $p > q$.
    Since $h_0 = 1$, we get a family of solutions
    $$h_i = \left(\frac{q}{p}\right)^i + A\left(1 - \left(\frac{q}{p}\right)^i\right).$$
    For a non-negative solution we must have $A \geq 0$.
    So the minimal nonnegative solution is $h_i = (\frac{q}{p})^i$.
- Suppose $p = q$.
  The recurrence relation has a general solution $h_i = A + Bi$.
  Again, $0 \leq h_i \leq 1$ forces $B = 0$. So $h_i = 1$, for all i.
  Thus, even in a fair casino, we are certain to end up broke.
  This apparent paradox is called **gamblers' ruin**.

Example: Birth-and-Death Chain

- Consider the following Markov chain.



For $i = 1, 2, \ldots$, we have $0 < p_i = 1 - q_i < 1$.

As in the preceding example, 0 is an absorbing state.

We wish to calculate the absorption probability starting from i.

Such a chain may serve as a model for the size of a population.

p_i is the probability of a birth before a death in a population of size i.

Then $h_i = \mathbb{P}_i(\text{hit } 0)$ is the extinction probability starting from i.

We write down the usual system of equations

$$\begin{aligned} h_0 &= 1, \\ h_i &= p_i h_{i+1} + q_i h_{i-1}, \; i = 1, 2, \ldots. \end{aligned}$$

This recurrence relation has variable coefficients.

Example: Birth-and-Death Chain (Cont'd)

- Take $h_i = p_i h_{i+1} + q_i h_{i-1}$.
  Rewrite as
  $$p_i h_i + q_i h_i = p_i h_{i+1} + q_i h_{i-1}.$$
  Consider $u_i = h_{i-1} - h_i$.
  Then
  $$p_i u_{i+1} = q_i u_i.$$
  So
  $$u_{i+1} = \left(\frac{q_i}{p_i}\right) u_i = \left(\frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1}\right) u_1 = \gamma_i u_1,$$
  where $\gamma_i := \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1}$.
  Then
  $$u_1 + \cdots + u_i = h_0 - h_i.$$

Example: Birth-and-Death Chain (Cont'd)

- We now have

$$h_i = 1 - A(\gamma_0 + \cdots + \gamma_{i-1}),$$

where $A = u_1$ and $\gamma_0 = 1$, with A still to be determined.

  - Suppose $\sum_{i=0}^{\infty} \gamma_i = \infty$.
    The restriction $0 \leq h_i \leq 1$ forces $A = 0$.
    So $h_i = 1$, for all i.
  - Suppose $\sum_{i=0}^{\infty} \gamma_i < \infty$.
    Then we can take $A > 0$ so long as

$$1 - A(\gamma_0 + \cdots + \gamma_{i-1}) \geq 0, \quad \text{for all } i.$$

    Thus, the minimal non-negative solution occurs when $A = \frac{1}{\sum_{i=0}^{\infty} \gamma_i}$.
    Then

$$h_i = \frac{\sum_{j=i}^{\infty} \gamma_j}{\sum_{j=0}^{\infty} \gamma_j}.$$

    In this case, for $i = 1, 2, \ldots$, we have $h_i < 1$.
    So the population survives with positive probability.

Mean Hitting Times

- Recall that

$$k_i^A = \mathbb{E}_i(H^A),$$

where H^A is the first time $(X_n)_{n \geq 0}$ hits A.

- We use the notation 1_B for the indicator function of B.

$$1_B(i) = \begin{cases} 1, & \text{if } i \in B, \\ 0, & \text{if } i \notin B. \end{cases}$$

Example: $1_{X_1 = j}$ is:

- Equal to 1 if $X_1 = j$;
- Equal to 0, otherwise.

Computing Mean Hitting Times

Theorem

The vector of mean hitting times $k^A = (k_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations

$$\begin{cases} k_i^A = 0, & \text{for } i \in A, \\ k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A, & \text{for } i \notin A. \end{cases}$$

- First we show that k^A satisfies the system.
  Suppose $X_0 = i \in A$. Then $H^A = 0$. So $k_i^A = 0$.
  Suppose $X_0 = i \notin A$. Then $H^A \geq 1$.
  By the Markov property, $\mathbb{E}_i(H^A | X_1 = j) = 1 + \mathbb{E}_j(H^A)$.

$$\begin{aligned} k_i^A &= \mathbb{E}_i(H^A) = \sum_{j \in I} \mathbb{E}_i(H^A 1_{X_1 = j}) \\ &= \sum_{j \in I} \mathbb{E}_i(H^A | X_1 = j) \mathbb{P}_i(X_1 = j) \\ &= 1 + \sum_{j \notin A} p_{ij} k_j^A. \end{aligned}$$

Mean Hitting Times (Converse)

- Suppose, now, that $y = (y_i : i \in I)$ is a solution to the given system. Suppose $i \in A$. Then $k_i^A = y_i = 0$.
  Suppose $i \notin A$. Then

$$\begin{array}{rcl} y_i & = & 1 + \sum_{j \notin A} p_{ij} y_j \\ & = & 1 + \sum_{j \notin A} p_{ij} (1 + \sum_{k \notin A} p_{jk} y_k) \\ & = & \mathbb{P}_i(H^A \geq 1) + \mathbb{P}_i(H^A \geq 2) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k. \end{array}$$

  By repeated substitution for y, we get after n steps

$$y_i = \mathbb{P}_i(H^A \geq 1) + \cdots + \mathbb{P}_i(H^A \geq n) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}.$$

  So, if y is non-negative, $y_i \geq \mathbb{P}_i(H^A \geq 1) + \cdots + \mathbb{P}_i(H^A \geq n)$.
  Letting $n \to \infty$,

$$y_i \geq \sum_{n=1}^{\infty} \mathbb{P}_i(H^A \geq n) = \mathbb{E}_i(H^A) = k_i^A.$$

Subsection 5

Strong Markov Property

Stopping Times

- Let $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ be a random variable.
- T is called a **stopping time** if the event $\{T = n\}$ depends only on $X_0, X_1, \ldots, X_n$, for $n = 0, 1, 2, \ldots$.

  Examples:

  (a) The **first passage time** $T_j = \inf\{n \geq 1 : X_n = j\}$ is a stopping time.
  We have $\{T_j = n\} = \{X_1 \neq j, \ldots, X_{n-1} \neq j, X_n = j\}$.

  (b) The first hitting time H^A is a stopping time.
  We have $\{H^A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$.

  (c) The **last exit time** $L^A = \sup\{n \geq 0 : X_n \in A\}$ is not in general a stopping time because the event $\{L^A = n\}$ depends on whether $(X_{n+m})_{m \geq 1}$ visits A or not.

Introducing the Strong Markov Property

- We shall show that the Markov Property holds at stopping times.
- The essential feature is that if:
  - T is a stopping time;
  - $B \subseteq \Omega$ is determined by $X_0, X_1, \ldots, X_T$;

  Then $B \cap \{T = m\}$ is determined by $X_0, X_1, \ldots, X_m$, for all $m = 0, 1, 2, \ldots$.

The Strong Markov Property

Theorem (Strong Markov Property)

Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and let T be a stopping time of $(X_n)_{n\geq 0}$. Then, conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \ldots, X_T$.

- Suppose B is an event determined by $X_0, X_1, \ldots, X_T$.

  Then $B \cap \{T = m\}$ is determined by $X_0, X_1, \ldots, X_m$.

  So, by the Markov Property at time m,

$$\begin{aligned} &\mathbb{P}(\{X_T = j_0, X_{T+1} = j_1, \ldots, X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}) \\ &= \mathbb{P}_i(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n)\mathbb{P}(B \cap \{T = m\} \cap \{X_T = i\}), \end{aligned}$$

  where we have used the condition $T = m$ to replace m by T.

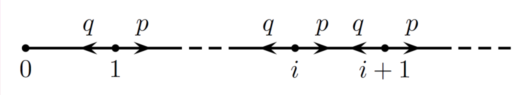
Strong Markov Property (Cont'd)

- We compute

$$\begin{aligned}
&\mathbb{P}(\{X_T = j_0, X_{T+1} = j_1, \ldots, X_{T+n} = j_n\} \cap B | T < \infty, X_T = i) \\
&= \frac{\mathbb{P}(\{X_T=j_0, X_{T+1}=j_1, \ldots, X_{T+n}=j_n\} \cap B \cap \{T<\infty, X_T=i\})}{\mathbb{P}(T<\infty, X_T=i)} \\
&= \frac{\sum_{m=0}^{\infty} \mathbb{P}(\{X_T=j_0, X_{T+1}=j_1, \ldots, X_{T+n}=j_n\} \cap B \cap \{T=m, X_T=i\})}{\mathbb{P}(T<\infty, X_T=i)} \\
&= \frac{\sum_{m=0}^{\infty} \mathbb{P}_i(\{X_T=j_0, X_{T+1}=j_1, \ldots, X_{T+n}=j_n\}) \mathbb{P}(B \cap \{T=m\} \cap \{X_T=i\})}{\mathbb{P}(T<\infty, X_T=i)} \\
&= \frac{\mathbb{P}_i(\{X_T=j_0, X_{T+1}=j_1, \ldots, X_{T+n}=j_n\}) \sum_{m=0}^{\infty} \mathbb{P}(B \cap \{T=m\} \cap \{X_T=i\})}{\mathbb{P}(T<\infty, X_T=i)} \\
&= \frac{\mathbb{P}_i(\{X_T=j_0, X_{T+1}=j_1, \ldots, X_{T+n}=j_n\}) \mathbb{P}(B \cap \{T<\infty\} \cap \{X_T=i\})}{\mathbb{P}(T<\infty, X_T=i)} \\
&= \mathbb{P}_i(\{X_T = j_0, X_{T+1} = j_1, \ldots, X_{T+n} = j_n\}) \mathbb{P}(B | T < \infty, X_T = i).
\end{aligned}$$

Example

- Consider the Markov chain $(X_n)_{n\geq 0}$ shown below.



Here, $0 < p = 1 - q < 1$.

We know from a previous example the probability of hitting 0 starting from 1.

We obtain the complete distribution of the time to hit 0 starting from 1 in terms of its probability generating function.

Set $H_j = \inf\{n \geq 0 : X_n = j\}$.

For $0 \leq s < 1$, let

$$\phi(s) = \mathbb{E}_1(s^{H_0}) = \sum_{n<\infty} s^n \mathbb{P}_1(H_0 = n).$$

Example (Cont'd)

- Suppose we start at 2.
  Apply the Strong Markov Property at H_1.
  Denote by $\widetilde{H}_0$ the time taken after H_1 to get to 0.
  - It is independent of H_1;
  - It has the (unconditioned) distribution of H_1.

  So, under $\mathbb{P}_2$, conditional on $H_1 < \infty$, we have

$$H_0 = H_1 + \widetilde{H}_0.$$

  Now we get

$$\begin{aligned}
\mathbb{E}_2(s^{H_0}) &= \mathbb{E}_2(s^{H_1} | H_1 < \infty)\mathbb{E}_2(s^{\widetilde{H}_0} | H_1 < \infty)\mathbb{P}_2(H_1 < \infty) \\
&= \mathbb{E}_2(s^{H_1} 1_{H_1 < \infty})\mathbb{E}_2(s^{\widetilde{H}_0} | H_1 < \infty) \\
&= \mathbb{E}_2(s^{H_1})^2 \\
&= \phi(s)^2.
\end{aligned}$$

Example (Cont'd)

- Next we use the Markov Property at time 1, conditional on $X_1 = 2$.

  Let $\overline{H}_0$ be the time taken after time 1 to get to 0.

  It has the same distribution as H_0 does under $\mathbb{P}_2$.

  Moreover, we have

  $$H_0 = 1 + \overline{H}_0.$$

  So we get

  $$\begin{array}{rcl} \phi(s) & = & \mathbb{E}_1(s^{H_0}) \\ & = & p\mathbb{E}_1(s^{H_0}|X_1 = 2) + q\mathbb{E}_1(s^{H_0}|X_1 = 0) \\ & = & p\mathbb{E}_1(s^{1+\overline{H}_0}|X_1 = 2) + q\mathbb{E}_1(s|X_1 = 0) \\ & = & ps\mathbb{E}_2(s^{H_0}) + qs \\ & = & ps\phi(s)^2 + qs. \end{array}$$

  Thus $\phi = \phi(s)$ satisfies $ps\phi^2 - \phi + qs = 0$.

Example (Cont'd)

- We found that $\phi = \phi(s)$ satisfies $ps\phi^2 - \phi + qs = 0$.
  So $\phi = \frac{1 \pm \sqrt{1-4pqs^2}}{2ps}$.
  But $\phi(0) \leq 1$ and ϕ is continuous.
  So we are forced to take the negative root at $s = 0$ and stick with it for all $0 \leq s < 1$.
  To recover the distribution of H_0 we expand the square-root as a power series:

$$\begin{array}{rcl}\phi(s) & = & \frac{1-\sqrt{1-4pqs^2}}{2ps} \\ & = & \frac{1}{2ps}[1-(1+\frac{1}{2}(-4pqs^2)+\frac{1}{2}(-\frac{1}{2})\frac{(-4pqs^2)^2}{2!}+\cdots)] \\ & = & qs + pq^2s^3 + \cdots \\ & = & s\mathbb{P}_1(H_0=1) + s^2\mathbb{P}_1(H_0=2) + s^3\mathbb{P}_1(H_0=3) + \cdots .\end{array}$$

  The first few probabilities $\mathbb{P}_1(H_0 = 1)$, $\mathbb{P}_1(H_0 = 2)$, ... are readily checked from first principles.

Example (Cont'd)

- We found $\phi(s) = s\mathbb{P}_1(H_0 = 1) + s^2\mathbb{P}_1(H_0 = 2) + s^3\mathbb{P}_1(H_0 = 3) + \cdots$.

  On letting $s \nearrow 1$, we have $\phi(s) \to \mathbb{P}_1(H_0 < \infty)$.

  So

$$\begin{aligned} \mathbb{P}_1(H_0 < \infty) &= \frac{1-\sqrt{1-4pq}}{2p} \\ &\overset{q=1-p}{=} \frac{1-|2q-1|}{2p} \\ &= \begin{cases} 1, & \text{if } p \leq q, \\ \frac{q}{p}, & \text{if } p > q. \end{cases} \end{aligned}$$

  For the mean hitting time, $\mathbb{E}_1(H_0) = \lim_{s \nearrow 1} \phi'(s)$.

  It is only worth considering the case $p \leq q$, where the mean hitting time has a chance of being finite.

  Differentiate $ps\phi^2 - \phi + qs = 0$ to obtain $2ps\phi\phi' + p\phi^2 - \phi' + q = 0$.

  So $\phi'(s) = \frac{p\phi(s)^2+q}{1-2ps\phi(s)} \overset{s \nearrow 1}{\to} \frac{1}{1-2p} = \frac{1}{q-p}$.

Example

- We consider an application of the Strong Markov Property to a Markov chain $(X_n)_{n\geq 0}$ observed only at certain times.

  Suppose that J is some subset of the state-space I.

  Suppose we observe the chain only when it takes values in J.

  The resulting process $(Y_m)_{m\geq 0}$ may be obtained formally by setting $Y_m = X_{T_m}$, where

$$\begin{aligned} T_0 &= \inf\{n \geq 0 : X_n \in J\}; \\ T_{m+1} &= \inf\{n > T_m : X_n \in J\}, \quad m = 0, 1, 2, \ldots. \end{aligned}$$

  Let us assume that $\mathbb{P}(T_m < \infty) = 1$, for all m.

  For each m, T_m, the time of the m-th visit to J, is a stopping time.

Example (Cont'd)

- Let, for $j \in J$, the vector $(h_i^j : i \in I)$ be the minimal non-negative solution to
$$h_i^j = p_{ij} + \sum_{k \notin J} p_{ik} h_k^j;$$
Set, for $i, j \in J$, $\overline{p}_{ij} = h_i^j$.
By the Strong Markov Property, for $i_1, \ldots, i_{m+1} \in J$,
$$\begin{aligned} & \mathbb{P}(Y_{m+1} = i_{m+1} | Y_0 = i_1, \ldots, Y_m = i_m) \\ & = \mathbb{P}(X_{T_{m+1}} = i_{m+1} | X_{T_0} = i_1, \ldots, X_{T_m} = i_m) \\ & = \mathbb{P}_{i_m}(X_{T_1} = i_{m+1}) = \overline{p}_{i_m i_{m+1}}. \end{aligned}$$
Thus $(Y_m)_{m \geq 0}$ is a Markov chain on J with transition matrix P.

Example

- A second example of a similar type arises if we observe the original chain $(X_n)_{n\geq 0}$ only when it moves.
  The resulting process $(Z_m)_{m\geq 0}$ is given by $Z_m = X_{S_m}$, where $S_0 = 0$ and for $m = 0, 1, 2, \ldots$,
  $$S_{m+1} = \inf\{n \in S_m : X_n \neq X_{S_m}\}.$$
  Let us assume there are no absorbing states.
  Then the random times S_m for $m \geq 0$ are stopping times.
  By the Strong Markov Property,
  $$\begin{aligned}&\mathbb{P}(Z_{m+1} = i_{m+1} | Z_0 = i_1, \ldots, Z_m = i_m)\\ &= \mathbb{P}(X_{S_{m+1}} = i_{m+1} | X_{S_0} = i_1, \ldots, X_{S_m} = i_m)\\ &= \mathbb{P}_{i_m}(X_{S_1} = i_{m+1}) = \widetilde{p}_{i_m i_{m+1}},\end{aligned}$$
  where $\widetilde{p}_{ii} = 0$ and, for $i \neq j$, $\widetilde{p}_{ij} = \frac{p_{ij}}{\sum_{k\neq i} p_{ik}}$.
  Thus $(Z_m)_{m\geq 0}$ is a Markov chain on I with transition matrix $\widetilde{P}$.

Subsection 6

Recurrence and Transience

Recurrent and Transient States

- Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P.
- We say that a state i is **recurrent** if
$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$
- We say that i is **transient** if
$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0.$$
- A recurrent state is one to which you keep coming back.
- A transient state is one which you eventually leave for ever.
- We will show that every state is either recurrent or transient.

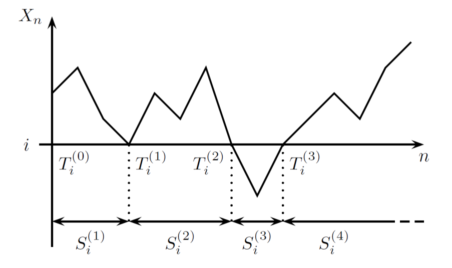
Passage Times

- The **first passage time** to state i is the random variable T_i defined by
$$T_i(\omega) = \inf\{n \geq 1 : X_n(\omega) = i\},$$
where $\inf \emptyset = \infty$.
- We now define inductively the r-**th passage time** $T_i^{(r)}$ to state i by
$$T_i^{(0)}(\omega) = 0;$$
$$T_i^{(1)}(\omega) = T_i(\omega);$$
$$T_i^{(r+1)}(\omega) = \inf\{n \geq T_i^{(r)}(\omega) + 1 : X_n(\omega) = i\}, \ r = 0, 1, \ldots.$$

Length of Excursion

- The length of the r-th excursion to i is

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)}, & \text{if } T_i^{(r-1)} < \infty, \\ 0, & \text{otherwise.} \end{cases}$$



Excursion Lengths Given Passage Times

Lemma

For $r = 2, 3, \ldots$, conditional on $T_i^{(r-1)} < \infty$, $S_i^{(r)}$ is independent of $\{X_m : m \leq T_i^{(r-1)}\}$ and $\mathbb{P}(S_i^{(r)} = n | T_i^{(r-1)} < \infty) = \mathbb{P}_i(T_i = n)$.

- Apply the strong Markov property at the stopping time $T = T_i^{(r-1)}$.

  It is automatic that $X_T = i$ on $T < \infty$.

  So, conditional on $T < \infty$:

  - $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P);
  - Independent of $X_0, X_1, \ldots, X_T$.

  But

  $$S_i^{(r)} = \inf\{n \geq 1 : X_{T+n} = i\}.$$

  So $S_i^{(r)}$ is the first passage time of $(X_{T+n})_{n \geq 0}$ to state i.

Number of Visits and Return Probabilities

- Recall that the indicator function $1_{\{X_1=j\}}$ is the random variable equal to 1 if $X_1 = j$ and 0 otherwise.
- We introduce the **number of visits** V_i to i.
- It may be written in terms of indicator functions as

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n=i\}}.$$

- Note that

$$\begin{array}{rcl} \mathbb{E}_i(V_i) & = & \mathbb{E}_i \sum_{n=0}^{\infty} 1_{\{X_n=i\}} \\ & = & \sum_{n=0}^{\infty} \mathbb{E}_i(1_{\{X_n=i\}}) \\ & = & \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) \\ & = & \sum_{n=0}^{\infty} p_{ii}^{(n)}. \end{array}$$

- Define the **return probability** $f_i = \mathbb{P}_i(T_i < \infty)$.

Number of Visits in terms of Return Probabilities

Lemma

For $r = 0, 1, 2, \ldots$, we have $\mathbb{P}_i(V_i > r) = f_i^r$.

- Observe that if $X_0 = i$, then $\{V_i > r\} = \{T_i^{(r)} < \infty\}$.

  When $r = 0$ the result is true.

  Suppose inductively that it is true for r.

  Then

$$\begin{array}{rcl}
\mathbb{P}_i(V_i > r+1) & = & \mathbb{P}_i(T_i^{(r+1)} < \infty) \\
& = & \mathbb{P}_i(T_i^{(r)} < \infty \text{ and } S_i^{(r+1)} < \infty) \\
& = & \mathbb{P}_i(S_i^{(r+1)} < \infty | T_i^{(r)} < \infty)\mathbb{P}_i(T_i^{(r)} < \infty) \\
& \overset{\text{prec. lem.}}{=} & f_i f_i^r \\
& = & f_i^{r+1}.
\end{array}$$

Expectation of Nonnegative Integer Random Variable

- Recall that one can compute the expectation of a non-negative integer-valued random variable as follows:

$$\begin{aligned}\mathbb{E}(V) &= \sum_{v=1}^{\infty} v\mathbb{P}(V=v) \\ &= \sum_{v=1}^{\infty} \sum_{r=0}^{v-1} \mathbb{P}(V=v) \\ &= \sum_{r=0}^{\infty} \sum_{v=r+1}^{\infty} \mathbb{P}(V=v) \\ &= \sum_{r=0}^{\infty} \mathbb{P}(V>r).\end{aligned}$$

Criterion for Recurrence or Transience

Theorem

The following dichotomy holds:

(i) if $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$;

(ii) if $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.

In particular, every state is either transient or recurrent.

- If $\mathbb{P}_i(T_i < \infty) = 1$, then, by the preceding lemma,

$$\mathbb{P}_i(V_i = \infty) = \lim_{r \to \infty} \mathbb{P}_i(V_i > r) = 1.$$

So i is recurrent and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i(V_i) = \infty.$$

Criterion for Recurrence or Transience (Cont'd)

- On the other hand, suppose $f_i = \mathbb{P}_i(T_i < \infty) < 1$.

  Then by the preceding lemma

$$\begin{array}{rcl} \sum_{n=0}^{\infty} p_{ii}^{(n)} & = & \mathbb{E}_i(V_i) \\ & = & \sum_{r=0}^{\infty} \mathbb{P}_i(V_i > r) \\ & = & \sum_{r=0}^{\infty} f_i^r \\ & = & \frac{1}{1-f_i} \\ & < & \infty. \end{array}$$

  So $\mathbb{P}_i(V_i = \infty) = 0$ and i is transient.

Class Property of Recurrence and Transience

Theorem

Let C be a communicating class. Then either all states in C are transient or all are recurrent.

- Take any pair of states $i, j \in C$ and suppose that i is transient. By hypothesis, there exist $n, m \geq 0$ with $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. Moreover, for all $r \geq 0$,

$$p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}.$$

  So, by the preceding theorem,

$$\sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(n+r+m)} < \infty.$$

  Hence j is also transient.
- As a result, we may speak of a **recurrent** or **transient class**.

Closure of Recurrent Classes

Theorem

Every recurrent class is closed.

- Let C be a class which is not closed.
  Then there exist $i \in C$, $j \notin C$ and $m \geq 1$, with $\mathbb{P}_i(X_m = j) > 0$.
  But we have
  $$\mathbb{P}_i(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}) = 0.$$
  It follows that
  $$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) < 1.$$
  So i is not recurrent.
  Hence, neither is C.

A Partial Converse

Theorem

Every finite closed class is recurrent.

- Suppose C is closed and finite and that $(X_n)_{n\geq 0}$ starts in C.

  Then for some $i \in C$ we have

$$\begin{aligned} 0 &< \mathbb{P}(X_n = i \text{ for infinitely many } n) \\ &= \mathbb{P}(X_n = i \text{ for some } n)\mathbb{P}_i(X_n = i \text{ for infinitely many } n). \\ & \quad \text{(Strong Markov Property)} \end{aligned}$$

  This shows that i is not transient.

  So C is recurrent by previous theorems.

Property of Irreducible and Recurrent Chains

- Remember that *irreducibility* means that the chain can get from any state to any other, with positive probability.

Theorem

Suppose P is irreducible and recurrent. Then for all $j \in I$,

$$\mathbb{P}(T_j < \infty) = 1.$$

- By the Markov Property we have

$$\mathbb{P}(T_j < \infty) = \sum_{i \in I} \mathbb{P}(X_0 = i)\mathbb{P}_i(T_j < \infty).$$

So it suffices to show that, for all $i \in I$,

$$\mathbb{P}_i(T_j < \infty) = 1.$$

Property of Irreducible and Recurrent Chains (Cont'd)

- Choose m with $p_{ji}^{(m)} > 0$.

  By a previous theorem, we have

$$\begin{array}{rcl} 1 & = & \mathbb{P}_j(X_n = j \text{ for infinitely many } n) \\ & = & \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1) \\ & = & \sum_{k \in I} \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1 | X_m = k) \mathbb{P}_j(X_m = k) \\ & \stackrel{\text{Markov}}{=} & \sum_{k \in I} \mathbb{P}_k(T_j < \infty) p_{jk}^{(m)}. \end{array}$$

  But $\sum_{k \in I} p_{jk}^{(m)} = 1$.

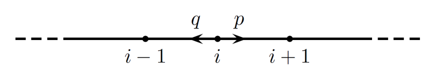
  So we must have $\mathbb{P}_i(T_j < \infty) = 1$, for all $i \in I$.

Subsection 7

Recurrence and Transience of Random Walks

Example: Simple Random Walk on $\mathbb{Z}$

- The simple random walk on $\mathbb{Z}$ has the following diagram.



As usual, we have $0 < p = 1 - q < 1$.
Suppose we start at 0.
  - It is clear that we cannot return to 0 after an odd number of steps. So $p_{00}^{(2n+1)} = 0$, for all n.
  - Any given sequence of steps of length $2n$ from 0 to 0 occurs with probability $p^n q^n$, there being n steps up and n steps down.
    The number of such sequences is the number of ways of choosing the n steps up from $2n$. Thus, $p_{00}^{(2n)} = \binom{2n}{n} p^n q^n$.

Stirling's formula provides a good approximation to $n!$ for large n,

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \text{ as } n \to \infty,$$

where $a_n \sim b_n$ means $\frac{a_n}{b_n} \to 1$.

Example (Cont'd)

- For the n-step transition probabilities we obtain

$$p_{00}^{(2n)} = \frac{(2n)!}{(n!)^2}(pq)^n \sim \frac{(4pq)^n}{A\sqrt{n/2}} \text{as } n \to \infty.$$

  - In the symmetric case $p = q = \frac{1}{2}$. So $4pq = 1$. Then, for some N and all $n \geq N$, we have $p_{00}^{(2n)} \geq \frac{1}{2A\sqrt{n}}$. So

$$\sum_{n=N}^{\infty} p_{00}^{(2n)} \geq \frac{1}{2A} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty.$$

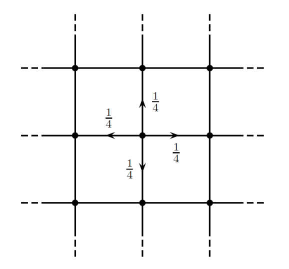
    This shows that the random walk is recurrent.
  - If $p \neq q$, then $4pq = r < 1$. So by a similar argument, for some N

$$\sum_{n=N}^{\infty} p_{00}^{(n)} \leq \frac{1}{A} \sum_{n=N}^{\infty} r^n < \infty.$$

    This shows that the random walk is transient.

Example: Simple Symmetric Random Walk on $\mathbb{Z}^2$

- The simple symmetric random walk on $\mathbb{Z}^2$ is shown below.



- The transition probabilities are given by

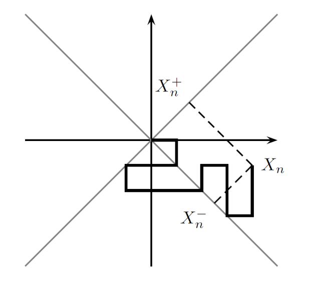
$$p_{ij} = \begin{cases} \frac{1}{4}, & \text{if } |i-j| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example: Simple Symmetric Random Walk on $\mathbb{Z}^2$ (Cont'd)

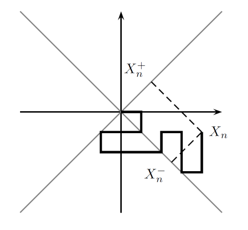
- Suppose we start at 0.

  We call the walk X_n.

  We write:
  - X_n^+ for the orthogonal projection of X_n on $y = x$;
  - X_n^- for the orthogonal projection of X_n on $y = -x$.



Example: Simple Symmetric Random Walk on $\mathbb{Z}^2$ (Cont'd)



- X_n^+ and X_n^- are independent symmetric random walks on $2^{-1/2}\mathbb{Z}$.
  Moreover, $X_n = 0$ if and only if $X_n^+ = 0 = X_n^-$.
  This makes it clear that for X_n we have (using Stirling's formula)

$$p_{00}^{(2n)} = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2 \sim \frac{2}{A^2 n} \text{ as } n \to \infty.$$

  Then $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$ by comparison with $\sum_{n=1}^{\infty} \frac{1}{n}$.
  So the walk is recurrent.

Example: Simple Symmetric Random Walk on $\mathbb{Z}^3$

- The transition probabilities of the simple symmetric random walk on $\mathbb{Z}^3$ are given by
$$p_{ij} = \begin{cases} \frac{1}{6}, & \text{if } |i-j| = 1, \\ 0, & \text{otherwise.} \end{cases}$$
Thus, the chain jumps to each of its nearest neighbors with equal probability.

  Suppose we start at 0.

  We can only return to 0 after an even number $2n$ of steps.

  Of these $2n$ steps there must be i up, i down, j north, j south, k east and k west for some $i, j, k \geq 0$, with
$$i + j + k = n.$$

Example: Simple Symmetric Random Walk on $\mathbb{Z}^3$ (Cont'd)

- By counting the ways in which this can be done, we obtain

$$\begin{aligned} p_{00}^{(2n)} &= \sum_{i,j,k\geq 0 i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} \left(\frac{1}{6}\right)^{2n} \\ &= \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k}^2 \left(\frac{1}{3}\right)^{2n}. \end{aligned}$$

The expression $\sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k}(\frac{1}{3})^n$ is the total probability of all the ways of placing n balls randomly into three boxes.

So we have

$$\sum_{\substack{i,j,k\geq 0\\ i+j+k=n}} \binom{n}{i\ j\ k} \left(\frac{1}{3}\right)^n = 1.$$

Example: Simple Symmetric Random Walk on $\mathbb{Z}^3$ (Cont'd)

- For the case where $n = 3m$, we have, for all i, j, k,

$$\binom{n}{i\ j\ k} = \frac{n!}{i!j!k!} \leq \binom{n}{m\ m\ m}.$$

So, using Stirling's formula,

$$p_{00}^{(2n)} \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m\ m\ m}\left(\frac{1}{3}\right)^{n} \sim \frac{1}{2A^3}\left(\frac{6}{n}\right)^{3/2} \text{ as } n \to \infty.$$

Hence, $\sum_{m=0}^{\infty} p_{00}^{(6m)} < \infty$, by comparison with $\sum_{n=0}^{\infty} n^{-3/2}$.

But we have, for all m:

- $p_{00}^{(6m)} \geq (\frac{1}{6})^2 p_{00}^{(6m-2)}$;
- $p_{00}^{(6m)} \geq (\frac{1}{6})^4 p_{00}^{(6m-4)}$.

So we must have $\sum_{n=0}^{\infty} p_{00}^{(n)} < \infty$. So the walk is transient.

Subsection 8

Invariant Distributions

Invariant Distributions

- Recall that a **measure** λ is any row vector $(\lambda_i : i \in I)$ with non-negative entries.
- We say λ is **invariant** if $\lambda P = \lambda$.
- Alternative terms are **equilibrium** and **stationary**.

The Stationary Property

- The first result explains the term *stationary*.

Theorem

Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and suppose that λ is invariant for P. Then $(X_{m+n})_{n\geq 0}$ is also Markov(λ, P).

- By a previous theorem, $\mathbb{P}(X_m = i) = (\lambda P^m)_i = \lambda_i$, for all i.
  Moreover, conditional on $X_{m+n} = i$:
  - X_{m+n+1} is independent of $X_m, X_{m+1}, \ldots, X_{m+n}$;
  - It has distribution $(p_{ij} : j \in I)$.

The Equilibrium Property

- The next result explains the term *equilibrium*.

Theorem

Let I be finite. Suppose that, for some $i \in I$,

$$p_{ij}^{(n)} \to \pi_j \text{ as } n \to \infty, \text{ for all } j \in I.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

- We have

$$\sum_{j \in I} \pi_j = \sum_{j \in I} \lim_{n \to \infty} p_{ij}^{(n)} = \lim_{n \to \infty} \sum_{j \in I} p_{ij}^{(n)} = 1.$$

Here, finiteness of I justifies interchange of summation and limit operations.

The Equilibrium Property (Cont'd)

- We saw that $\sum_{j \in I} \pi_j = 1$.

  We also have

$$\begin{aligned} \pi_j &= \lim_{n\to\infty} p_{ij}^{(n)} \\ &= \lim_{n\to\infty} \sum_{k\in I} p_{ik}^{(n)} p_{kj} \\ &= \sum_{k\in I} \lim_{n\to\infty} p_{ik}^{(n)} p_{kj} \\ &= \sum_{k\in I} \pi_k p_{kj}, \end{aligned}$$

  where, again, finiteness of I justifies interchange of summation and limit operations.

  Hence, π is an invariant distribution.

- Notice that for any of the random walks discussed in the preceding subsection, we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$, for all $i, j \in I$.

  The limit is certainly invariant, but it is not a distribution!

Example

- Consider the two-state Markov chain with transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

  Ignore the trivial cases $\alpha = \beta = 0$ and $\alpha = \beta = 1$.

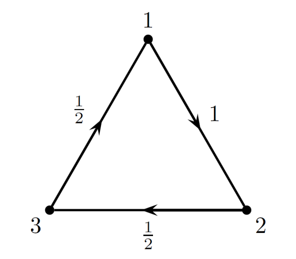
  By a previous example,

$$P^n \to \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} \text{ as } n \to \infty.$$

  So, by the preceding theorem, the distribution $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$ must be invariant.
- There are, of course, easier ways to discover this.

Example

- Consider the Markov chain $(X_n)_{n\geq 0}$ with the diagram shown.



Then

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Let $\pi = (\pi_1, \pi_2, \pi_3)$.

To find an invariant distribution we write down the components of the vector equation $\pi P = \pi$.

We have

$$\pi P = (\pi_1, \pi_2, \pi_3) \begin{pmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = \left(\frac{1}{2}\pi_3, \pi_1 + \frac{1}{2}\pi_2, \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3\right).$$

Example (Cont'd)

- So $\pi P = \pi$ gives
$$\pi_1 = \frac{1}{2}\pi_3,$$
$$\pi_2 = \pi_1 + \frac{1}{2}\pi_2,$$
$$\pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3.$$

  In terms of the chain:
  - The right sides give the probabilities for X_1, when X_0 has distribution π;
  - The equations require X_1 also to have distribution π.

  The equations are homogeneous so one of them is redundant.

  Thus, another equation is required to fix π uniquely,
$$\pi_1 + \pi_2 + \pi_3 = 1.$$

  Solving, we find that $\pi = (\frac{1}{5}, \frac{2}{5}, \frac{2}{5})$.

Invariant Distribution for Finite State Space

- For a finite state space I, the existence of an invariant row vector follows by linear algebra.

  The row sums of P are all 1.

  So the column vector of ones is an eigenvector with eigenvalue 1.

  So P must have a row eigenvector with eigenvalue 1.

Time Spent Between Visits

- Fix a state k.
- Consider, for each i, the **expected time spent in i between visits to k**,

$$\gamma_i^k = \mathbb{E}_k \sum_{n=0}^{T_k-1} 1_{\{X_n=i\}}.$$

- Here the sum of indicator functions serves to count the number of times n at which $X_n = i$ before the first passage time T_k.

Properties of Time Spent Between Visits

Theorem

Let P be irreducible and recurrent. Then:

(i) $\gamma_k^k = 1$;

(ii) $\gamma^k = (\gamma_i^k : i \in I)$ satisfies $\gamma^k P = \gamma^k$;

(iii) $0 < \gamma_i^k < \infty$, for all $i \in I$.

(i) This is obvious.

(ii) For $n = 1, 2, \ldots$, the event $\{n \leq T_k\}$ depends only on $X_0, X_1, \ldots, X_{n-1}$. So, by the Markov property at $n - 1$,

$$\mathbb{P}_k(X_{n-1} = i, X_n = j \text{ and } n \leq T_k) = \mathbb{P}_k(X_{n-1} = i \text{ and } n \leq T_k) p_{ij}.$$

Since P is recurrent, under $\mathbb{P}_k$, we have:

- $T_k < \infty$;
- $X_0 = X_{T_k} = k$ with probability one.

Properties of Time Spent Between Visits (Cont'd)

- Therefore,

$$\begin{aligned}
\gamma_j^k &= \mathbb{E}_k \sum_{n=1}^{T_k} 1_{\{X_n=j\}} \\
&= \mathbb{E}_k \sum_{n=1}^{\infty} 1_{\{X_n=j \text{ and } n\leq T_k\}} \\
&= \sum_{n=1}^{\infty} \mathbb{P}_k(X_n=j \text{ and } n\leq T_k) \\
&= \sum_{i\in I}\sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1}=i, X_n=j \text{ and } n\leq T_k) \\
&= \sum_{i\in I} p_{ij}\sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1}=i \text{ and } n\leq T_k) \\
&= \sum_{i\in I} p_{ij}\mathbb{E}_k\sum_{m=0}^{\infty} 1_{\{X_m=i \text{ and } m\leq T_k-1\}} \\
&= \sum_{i\in I} p_{ij}\mathbb{E}_k\sum_{m=0}^{T_k-1} 1_{\{X_m=i\}} \\
&= \sum_{i\in I} \gamma_i^k p_{ij}.
\end{aligned}$$

(iii) By hypothesis, P is irreducible. So, for each state i, there exist $n, m \geq 0$, with $p_{ik}^{(n)}, p_{ki}^{(m)} > 0$. Then, using Parts (i) and (ii), $\gamma_i^k \geq \gamma_k^k p_{ki}^{(m)} > 0$. And, also, $\gamma_i^k p_{ik}^{(n)} \leq \gamma_k^k = 1$.

Invariant Measures and Time Spent Between Visits

Theorem

Let P be irreducible and let λ be an invariant measure for P with $\lambda_k = 1$. Then $\lambda \geq \gamma^k$. If, in addition, P is recurrent, then $\lambda = \gamma^k$.

- For each $j \in I$, we have

$$\begin{aligned}
\lambda_j &= \sum_{i_1 \in I} \lambda_{i_1} p_{i_1 j} \\
&= \sum_{i_1 \neq k} \lambda_{i_1} p_{i_1 j} + p_{kj} \\
&= \sum_{i_1, i_2 \neq k} \lambda_{i_2} p_{i_2 i_1} p_{i_1 j} + (p_{kj} + \sum_{i_1 \neq k} p_{k i_1} p_{i_1 j}) \\
&\vdots \\
&= \sum_{i_1, \ldots, i_n \neq k} \lambda_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 j} \\
&\quad + (p_{kj} + \sum_{i_1 \neq k} p_{k i_1} p_{i_1 j} + \cdots + \sum_{i_1, \ldots, i_{n-1} \neq k} p_{k i_{n-1}} \cdots p_{i_2 i_1} p_{i_1 j}).
\end{aligned}$$

Invariant Measures and Time Between Visits (Cont'd)

- So for $j \neq k$, we obtain

$$\begin{array}{rcl} \lambda_j & \geq & \mathbb{P}_k(X_1 = j \text{ and } T_k \geq 1) + \mathbb{P}_k(X_2 = j \text{ and } T_k \geq 2) \\ & & + \cdots + \mathbb{P}_k(X_n = j \text{ and } T_k \geq n) \\ & \to & \gamma_j^k \text{ as } n \to \infty. \end{array}$$

  So $\lambda \geq \gamma^k$.

  If P is recurrent, then γ^k is invariant by the preceding theorem.

  So $\mu = \lambda - \gamma^k$ is also invariant and $\mu \geq 0$.

  Since P is irreducible, given $i \in I$, we have $p_{ik}^{(n)} > 0$, for some n.

  So

$$0 = \mu_k = \sum_{j \in I} \mu_j p_{jk}^{(n)} \geq \mu_i p_{ik}^{(n)}.$$

  We conclude $\mu_i = 0$.

Positive Recurrence and Null Recurrence

- Recall that a state i is **recurrent** if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

- We showed that this is equivalent to $\mathbb{P}_i(T_i < \infty) = 1$.
- If, in addition, the expected return time

$$m_i = \mathbb{E}_i(T_i)$$

  is finite, then we say i is **positive recurrent**.
- A recurrent state which fails to have this stronger property is called **null recurrent**.

Positive Recurrence in Irreducible Chains

Theorem

Let P be irreducible. Then the following are equivalent:

(i) Every state is positive recurrent;

(ii) Some state i is positive recurrent;

(iii) P has an invariant distribution, π say.

Moreover, when (iii) holds we have $m_i = \frac{1}{\pi_i}$, for all i.

(i)⇒(ii) is obvious.

(ii)⇒(iii) If i is positive recurrent, it is certainly recurrent.

So P is recurrent.

By a previous theorem, γ^i is then invariant.

But $\sum_{j \in I} \gamma_j^i = m_i < \infty$.

So $\pi_j = \frac{\gamma_j^i}{m_i}$ defines an invariant distribution.

Positive Recurrence in Irreducible Chains (Cont'd)

(iii)$\Rightarrow$(i) Take any state k.

Now P is irreducible and $\sum_{i\in I}\pi_i = 1$.

So we have $\pi_k = \sum_{i\in I}\pi_i p_{ik}^{(n)} > 0$, for some n.

Set

$$\lambda_i = \frac{\pi_i}{\pi_k}.$$

Then λ is an invariant measure with $\lambda_k = 1$.

So by the preceding theorem, $\lambda \geq \gamma^k$.

Hence,

$$m_k = \sum_{i\in I}\gamma_i^k \leq \sum_{i\in I}\frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty.$$

So k is positive recurrent.

To complete the proof we revisit the argument for (iii)$\Rightarrow$(i).

Now we know that P is recurrent.

Then $\lambda = \gamma^k$ and the preceding inequality is in fact an equality.

Example: Simple Symmetric Random Walk on $\mathbb{Z}$

- The simple symmetric random walk on $\mathbb{Z}$ is clearly irreducible.

  By a previous example, it is also recurrent.

  Consider the measure $\pi_i = 1$, for all i.

  Then
  $$\pi_i = \frac{1}{2}\pi_{i-1} + \frac{1}{2}\pi_{i+1}.$$

  So π is invariant.

  By a previous theorem, any invariant measure is a scalar multiple of π.

  But $\sum_{i\in\mathbb{Z}} \pi_i = \infty$.

  So there can be no invariant distribution.

  Thus, the walk is null recurrent, by the preceding theorem.

Example

- The existence of an invariant measure does not guarantee recurrence.

  Consider, the simple symmetric random walk on $\mathbb{Z}^3$.

  By a previous example, it is transient.

  It has invariant measure π given by $\pi_i = 1$, for all i.

Example

- Consider the asymmetric random walk on $\mathbb{Z}$ with transition probabilities

$$p_{i,i-1} = q < p = p_{i,i+1}.$$

In components, the invariant measure equation $\pi P = \pi$ reads

$$\pi_i = \pi_{i-1} p + \pi_{i+1} q.$$

This is a recurrence relation for π.

It has general solution

$$\pi_i = A + B \left(\frac{p}{q} \right)^i.$$

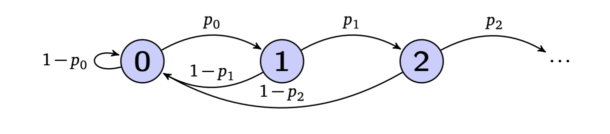
In this case, there is a two-parameter family of invariant measures.

This shows that uniqueness up to scalar multiples does not hold.

Example

- Consider a **success-run chain** on $\mathbb{Z}^+$, whose transition probabilities are given by

$$p_{i,i+1} = p_i, \quad p_{i0} = q_i = 1 - p_i.$$



Then the components of the invariant measure equation $\pi P = \pi$ read

$$\begin{aligned} \pi_0 &= \sum_{i=0}^{\infty} q_i \pi_i, \\ \pi_i &= p_{i-1}\pi_{i-1}, \quad \text{for } i \geq 1. \end{aligned}$$

Example (Cont'd)

- We have

$$\begin{array}{rcl}\pi_0 & = & \sum_{i=0}^{\infty} q_i \pi_i, \\ \pi_i & = & p_{i-1}\pi_{i-1}, \quad \text{for } i \geq 1.\end{array}$$

Suppose we choose p_i converging sufficiently rapidly to 1 so that

$$p = \prod_{i=0}^{\infty} p_i > 0.$$

Then for any invariant measure π we have

$$\pi_0 = \sum_{i=0}^{\infty} (1 - p_i) p_{i-1} \cdots p_0 \pi_0 = (1 - p)\pi_0.$$

This equation forces either $\pi_0 = 0$ or $\pi_0 = \infty$.

So there is no non-zero invariant measure.

Subsection 9

Convergence to Equilibrium

Limiting Behavior of n-Step Probabilities

- We saw that, if the state space is finite, and, for some i, the limit π_i of p_{ij}^n as $n \to \infty$ exists, for all j, then π must be an invariant distribution.
- But the limit does not always exist.

  Example: Consider the two-state chain with transition matrix

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

  Then $P^2 = I$.

  So $P^{2n} = I$ and $P^{2n+1} = P$, for all n.

  Thus $p_{ij}^{(n)}$ fails to converge for all i, j.

Aperiodic States

- We call a state i **aperiodic** if $p_{ii}^{(n)} > 0$, for all sufficiently large n.
- It is easy to show that i is aperiodic if and only if the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$ has no common divisor other than 1.

Lemma

Suppose P is irreducible and has an aperiodic state i. Then, for all states j and k, $p_{jk}^{(n)} > 0$ for all sufficiently large n. In particular, all states are aperiodic.

- By irreducibility, there exist $r, s \geq 0$, with $p_{ji}^{(r)}, p_{ik}^{(s)} > 0$. Then, for all sufficiently large n,

$$p_{jk}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0.$$

Convergence to Equilibrium

Theorem (Convergence to Equilibrium)

Let P be irreducible and aperiodic, and suppose that P has an invariant distribution π. Let λ be any distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then

$$\mathbb{P}(X_n = j) \to \pi_j \text{ as } n \to \infty, \text{ for all } j.$$

In particular, $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$, for all i, j.

- We use a coupling argument.
  Let $(Y_n)_{n\geq 0}$ be Markov(π, P) and independent of $(X_n)_{n\geq 0}$.
  Fix a reference state b and set

$$T = \inf\{n \geq 1 : X_n = Y_n = b\}.$$

Convergence to Equilibrium (Step 1)

- **Step 1**: We show $\mathbb{P}(T < \infty) = 1$.

  The process $W_n = (X_n, Y_n)$ is a Markov chain on $I \times I$ with:
  - Transition probabilities $\widetilde{p}_{(i,k)(j,\ell)} = p_{ij} p_{k\ell}$;
  - Initial distribution $\mu_{(i,k)} = \lambda_i \pi_k$.

  Since P is aperiodic, for all states i, j, k, ℓ, we have

  $$\widetilde{p}^{(n)}_{(i,k)(j,\ell)} = p^{(n)}_{ij} p^{(n)}_{k\ell} > 0,$$

  for all sufficiently large n. So $\widetilde{P}$ is irreducible.

  Also, $\widetilde{P}$ has an invariant distribution given by $\widetilde{\pi}_{(i,k)} = \pi_i \pi_k$.

  By a previous theorem, $\widetilde{P}$ is positive recurrent.

  But T is the first passage time of W_n to (b, b).

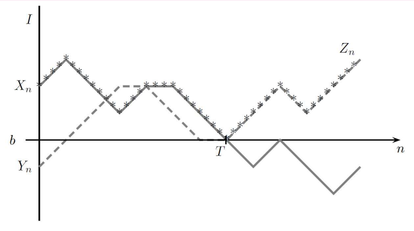
  By a previous theorem, $\mathbb{P}(T < \infty) = 1$.

Convergence to Equilibrium (Step 2)

- **Step 2**: Set

$$Z_n = \begin{cases} X_n, & \text{if } n < T \\ Y_n, & \text{if } n \geq T. \end{cases}$$

We show $(Z_n)_{n \geq 0}$ is Markov(λ, P).



The strong Markov property applies to $(W_n)_{n \geq 0}$ at time T.
So $(X_{T+n}, Y_{T+n})_{n \geq 0}$ is:

- Markov$(\delta_{(b,b)}, \widetilde{P})$;
- Independent of (X_0, Y_0), (X_1, Y_1), ..., (X_T, Y_T).

Convergence to Equilibrium (Step 2 Cont'd)

- By symmetry, we can replace the process $(X_{T+n}, Y_{T+n})_{n\geq 0}$ by $(Y_{T+n}, X_{T+n})_{n\geq 0}$.

  This is also:
  - Markov$(\delta_{(b,b)}, \widetilde{P})$;
  - Independent of (X_0, Y_0), (X_1, Y_1), …, (X_T, Y_T).

  Hence $W_n' = (Z_n, Z_n')$ is Markov$(\mu, \widetilde{P})$, where

$$Z_n' = \begin{cases} Y_n, & \text{if } n < T, \\ X_n, & \text{if } n \geq T. \end{cases}$$

  In particular, $(Z_n)_{n\geq 0}$ is Markov(λ, P).

Convergence to Equilibrium (Step 3)

- **Step 3**: We have

$$\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j \text{ and } n < T) + \mathbb{P}(Y_n = j \text{ and } n \geq T).$$

So

$$\begin{aligned} |\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - P(Y_n = j)| \\ &= |\mathbb{P}(X_n = j \text{ and } n < T) \\ &\qquad - \mathbb{P}(Y_n = j \text{ and } n < T)| \\ &\leq \mathbb{P}(n < T). \end{aligned}$$

The result follows since $\mathbb{P}(n < T) \to 0$ as $n \to \infty$.

Example: Non-Aperiodic Transitions

- To understand this proof one should see what goes wrong when P is not aperiodic.

  Example: Consider the two-state chain with transition matrix

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

  It has $(\frac{1}{2}, \frac{1}{2})$ as its unique invariant distribution.

  We start:
  - $(X_n)_{n \geq 0}$ from 0;
  - $(Y_n)_{n \geq 0}$ with equal probability from 0 or 1.

  Suppose $Y_0 = 1$.

  Because of periodicity, $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ will never meet.

  So, in this case, the proof fails.

Decomposition of the State Space

Theorem

Let P be irreducible. There is an integer $d \geq 1$ and a partition $I = C_0 \cup C_1 \cup \cdots \cup C_{d-1}$, such that (setting $C_{nd+r} = C_r$):

(i) $p_{ij}^{(n)} > 0$ only if $i \in C_r$ and $j \in C_{r+n}$, for some r;

(ii) $p_{ij}^{(nd)} > 0$ for all sufficiently large n, for all $i, j \in C_r$, for all r.

- Fix a state k and consider $S = \{n \geq 0 : p_{kk}^{(n)} > 0\}$.
  Choose $n_1, n_2 \in S$, with:
  - $n_1 < n_2$;
  - $d := n_2 - n_1$ is as small as possible.

  Define for $r = 0, \ldots, d-1$,

$$C_r = \{i \in I : p_{ki}^{(nd+r)} > 0 \text{ for some } n \geq 0\}.$$

  By irreducibility, $C_0 \cup \cdots \cup C_{d-1} = I$.

Decomposition of the State Space (Cont'd)

- Suppose, for some $r, s \in \{0, 1, \ldots, d-1\}$, we have:
  - $p_{ki}^{(nd+r)} > 0$;
  - $p_{ki}^{(nd+s)} > 0$.

  Choose $m \geq 0$ so that $p_{ik}^{(m)} > 0$.

  Then we have:
  - $p_{kk}^{(nd+r+m)} > 0$;
  - $p_{kk}^{(nd+s+m)} > 0$.

  So $r = s$ by minimality of d.

  Hence we have a partition.

Decomposition of the State Space (Part (i))

- Now we prove Part (i).
  Suppose $p_{ij}^{(n)} > 0$ and $i \in C_r$. Choose m so that $p_{ki}^{(md+r)} > 0$.
  Then $p_{kj}^{(md+r+n)} > 0$. So $j \in C_{r+n}$, as claimed.
  By taking $i = j = k$, we see that d must divide every element of S.
  In particular d must divide n_1.
  For $nd \geq n_1^2$, we can write

  $$nd = qn_1 + r,$$

  for integers $q \geq n_1$ and $0 \leq r \leq n_1 - 1$.
  Since d divides n_1, we then have $r = md$, for some integer m.
  Then $nd = (q - m)n_1 + mn_2$.
  Hence

  $$p_{kk}^{(nd)} \geq (p_{kk}^{(n_1)})^{q-m}(p_{kk}^{(n_2)})^m > 0.$$

  So $nd \in S$.

Decomposition of the State Space (Part (ii))

- Now we prove Part (ii).

  For $i, j \in C_r$, choose m_1 and m_2 so that:
  - $p_{ik}^{(m_1)} > 0$;
  - $p_{kj}^{(m_2)} > 0$.

  Then, if $nd \geq n_1^2$,

$$p_{ij}^{(m_1+nd+m_2)} \geq p_{ik}^{(m_1)} p_{kk}^{(nd)} p_{kj}^{(m_2)} > 0.$$

  But, by Part (i), $m_1 + m_2$ is then necessarily a multiple of d.

  This concludes the proof.
- We call d the **period** of P.
- The theorem shows, in particular, for all $i \in I$, that d is the greatest common divisor of the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$.

Description of Limiting Behavior for Irreducible Chains

Theorem

Let P be irreducible of period d and let $C_0, C_1, \ldots, C_{d-1}$ be the partition obtained in the preceding theorem. Let λ be a distribution with $\sum_{i \in C_0} \lambda_i = 1$. Suppose that $(X_n)_{n \geq 0}$ is Markov(λ, P). Then for $r = 0, 1, \ldots, d-1$ and $j \in C_r$ we have

$$\mathbb{P}(X_{nd+r} = j) \to \frac{d}{m_j} \text{ as } n \to \infty,$$

where m_j is the expected return time to j. In particular, for $i \in C_0$ and $j \in C_r$ we have

$$p_{ij}^{(nd+r)} \to \frac{d}{m_j} \text{ as } n \to \infty.$$

Limiting Behavior for Irreducible Chains (Step 1)

- **Step 1**: We reduce to the aperiodic case.
  Set $\nu = \lambda P^r$. By the preceding theorem, $\sum_{i \in C_r} \nu_i = 1$.
  Set $Y_n = X_{nd+r}$. Then $(Y_n)_{n \geq 0}$ is Markov(ν, P^d).
  By the preceding theorem, P^d is irreducible and aperiodic on C_r.
  For $j \in C_r$ the expected return time of $(Y_n)_{n \geq 0}$ to j is $\frac{m_j}{d}$.
  Assume the theorem holds in the aperiodic case.
  Then
  $$\mathbb{P}(X_{nd+r} = j) = \mathbb{P}(Y_n = j) \to \frac{d}{m_j} \text{ as } n \to \infty.$$
  So the theorem holds in general.

Limiting Behavior for Irreducible Chains (Step 2)

- **Step 2**: Assume that P is aperiodic.

  If P is positive recurrent, then

  $$\frac{1}{m_j} = \pi_j,$$

  where π is the unique invariant distribution.

  So the result follows from a previous theorem.

  Otherwise, $m_j = \infty$.

  Then we have to show that

  $$\mathbb{P}(X_n = j) \to 0 \text{ as } n \to \infty.$$

  If P is transient this is easy.

  So we are left with the null recurrent case.

Limiting Behavior for Irreducible Chains (Step 3)

- **Step 3**: Assume that P is aperiodic and null recurrent. Then

$$\sum_{k=0}^{\infty} \mathbb{P}_j(T_j > k) = \mathbb{E}_j(T_j) = \infty.$$

Given $\varepsilon > 0$, choose K so that

$$\sum_{k=0}^{K-1} \mathbb{P}_j(T_j > k) \geq \frac{2}{\varepsilon}.$$

Then, for $n \geq K - 1$,

$$\begin{array}{rcl} 1 & \geq & \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j \text{ and } X_m \neq j \text{ for } m = k+1, \ldots, n) \\ & = & \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j)\mathbb{P}_j(T_j > n-k) \\ & = & \sum_{k=0}^{K-1} \mathbb{P}(X_{n-k} = j)\mathbb{P}_j(T_j > k). \end{array}$$

So we must have $\mathbb{P}(X_{n-k} = j) \leq \frac{\varepsilon}{2}$, for some $k \in \{0, 1, \ldots, K-1\}$.

Limiting Behavior for Irreducible Chains (Step 3 Cont'd)

- Return now to the coupling argument used in a previous theorem.
  Let $(Y_n)_{n\geq 0}$ be Markov(μ, P), where μ is to be chosen later.
  Set $W_n = (X_n, Y_n)$.
  As before, aperiodicity of $(X_n)_{n\geq 0}$ ensures irreducibility of $(W_n)_{n\geq 0}$.
  Assume, first, $(W_n)_{n\geq 0}$ is transient.
  Take $\mu = \lambda$.
  We obtain
  $$\mathbb{P}(X_n = j)^2 = \mathbb{P}(W_n = (j, j)) \to 0.$$
  Assume then that $(W_n)_{n\geq 0}$ is recurrent.
  Then we have $\mathbb{P}(T < \infty) = 1$.
  The coupling argument shows that
  $$|\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \to 0 \text{ as } n \to \infty.$$

Limiting Behavior for Irreducible Chains (Step 3 Cont'd)

- Take $\mu = \lambda P^k$, for $k = 1, \ldots, K-1$.

  Then

  $$\mathbb{P}(Y_n = j) = \mathbb{P}(X_{n+k} = j).$$

  We can find N, such that for $n \geq N$ and $k = 1, \ldots, K-1$,

  $$|\mathbb{P}(X_n = j) - \mathbb{P}(X_{n+k} = j)| \leq \frac{\varepsilon}{2}.$$

  But for any n, we can find $k \in \{0, 1, \ldots, K-1\}$, such that

  $$\mathbb{P}(X_{n+k} = j) \leq \frac{\varepsilon}{2}.$$

  Hence, for $n \geq N$, $\mathbb{P}(X_n = j) \leq \varepsilon$.

  Since $\varepsilon > 0$ was arbitrary, we get $\mathbb{P}(X_n = j) \to 0$ as $n \to \infty$.

Subsection 10

Time Reversal

Introducing Time Reversal

- For Markov chains, the past and future are independent given the present.
- This property is symmetrical in time and suggests looking at Markov chains with time running backwards.
- On the other hand, convergence to equilibrium shows behavior which is asymmetrical in time.
  - A highly organized state such as a point mass decays to a disorganized one, the invariant distribution.
  - This is an example of entropy increasing.
- It suggests that if we want complete time-symmetry we must begin in equilibrium.
  - We show that a Markov chain in equilibrium, run backwards, is again a Markov chain.
  - The transition matrix may however be different.

Time Reversal of an Irreducible Markov Chain

Theorem

Let P be irreducible and have an invariant distribution π. Suppose that $(X_n)_{0\leq n\leq N}$ is Markov(π, P) and set $Y_n = X_{N-n}$. Then $(Y_n)_{0\leq n\leq N}$ is Markov$(\pi, \widehat{P})$, where $\widehat{P} = (\widehat{p}_{ij})$ is given by

$$\pi_j \widehat{p}_{ji} = \pi_i p_{ij}, \text{ for all } i, j,$$

and $\widehat{P}$ is also irreducible with invariant distribution π.

- First we check that $\widehat{P}$ is a stochastic matrix:

$$\sum_{i\in I} \widehat{p}_{ji} = \frac{1}{\pi_j} \sum_{i\in I} \pi_i p_{ij} = 1. \quad (\pi \text{ invariant for } P)$$

  Next we check that π is invariant for $\widehat{P}$:

$$\sum_{j\in I} \pi_j \widehat{p}_{ji} = \sum_{j\in I} \pi_i p_{ij} = \pi_i. \quad (P \text{ stochastic})$$

Time Reversal of an Irreducible Markov Chain (Cont'd)

- We have
$$\begin{aligned}&\mathbb{P}(Y_0 = i_0, Y_1 = i_1, \ldots, Y_N = i_N)\\&= \mathbb{P}(X_0 = i_N, X_1 = i_{N-1}, \ldots, X_N = i_0)\\&= \pi_{i_N} p_{i_N i_{N-1}} \cdots p_{i_1 i_0}\\&= \pi_i \widehat{p}_{i_0 i_1} \cdots \widehat{p}_{i_{N-1} i_N}.\end{aligned}$$
So, by a previous theorem, $(Y_n)_{0 \leq n \leq N}$ is Markov$(\pi, \widehat{P})$.
Since P is irreducible, for each pair of states i, j, there is a chain of states $i_1 = i$, $i_2, \ldots, i_{n-1}$, $i_n = j$, with $p_{i_1 i_2} \cdots p_{i_{n-1} i_n} > 0$.
Then
$$\widehat{p}_{i_n i_{n-1}} \cdots \widehat{p}_{i_2 i_1} = \frac{\pi_{i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}}{\pi_{i_n}} > 0.$$
So $\widehat{P}$ is also irreducible.
- The chain $(Y_n)_{0 \leq n \leq N}$ is called the **time-reversal** of $(X_n)_{0 \leq n \leq N}$.

Detailed Balance

- A stochastic matrix P and a measure λ are said to be in **detailed balance** if

$$\lambda_i p_{ij} = \lambda_j p_{ji}, \text{ for all } i, j.$$

- When a solution λ to the detailed balance equations exists, it is often easier to find by the detailed balance equations than by the equation $\lambda = \lambda P$.

Lemma

If P and λ are in detailed balance, then λ is invariant for P.

- We have

$$(\lambda P)_i = \sum_{j \in I} \lambda_j p_{ji} = \sum_{j \in I} \lambda_i p_{ij} = \lambda_i.$$

Reversible Markov Chains

- Let $(X_n)_{n\geq 0}$ be Markov(λ, P), with P irreducible.
- We say that $(X_n)_{n\geq 0}$ is **reversible** if, for all $N \geq 1$, $(X_{N-n})_{0\leq n\leq N}$ is also Markov(λ, P).

Theorem

Let P be an irreducible stochastic matrix and let λ be a distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then the following are equivalent:

(a) $(X_n)_{n\geq 0}$ is reversible;

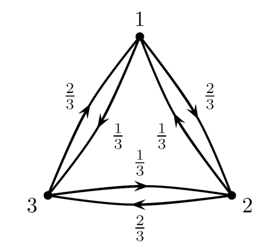
(b) P and λ are in detailed balance.

- Both (a) and (b) imply that λ is invariant for P.

  Then both (a) and (b) are equivalent to the statement that $\widehat{P} = P$ in the preceding theorem.

Example: A Non-Reversible Markov Chain

- Consider the Markov chain with diagram as on the right.
  The transition matrix is



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

and $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is invariant.

Hence $\widehat{P} = P^T$, the transpose of P.

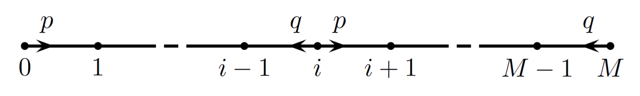
But P is not symmetric, so $P \neq \widehat{P}$.

Thus, this chain is not reversible.

A patient observer would see the chain move clockwise in the long run. Under time-reversal the clock would run backwards!

Example

- Consider the following Markov chain, where $0 < p = 1 - q < 1$.



The non-zero detailed balance equations read

$$\lambda_i p_{i,i+1} = \lambda_{i+1} p_{i+1,i}, \quad i = 0, 1, \ldots, M - 1.$$

So a solution is given by

$$\lambda = \left(\left(\frac{p}{q} \right)^i : i = 0, \ldots, M \right).$$

Normalized, this gives a distribution in detailed balance with P.

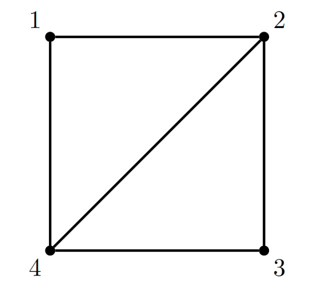
Hence, by the theorem, this chain is reversible.

Example (Comments)

- Suppose p were much larger than q.
- Then, one might argue that the chain would tend to move to the right and its time-reversal to the left.
- However, this ignores the fact that we reverse the chain in equilibrium.
- In this case, the equilibrium would be heavily concentrated near M.
- So the chain would spend most of its time near M, making occasional brief forays to the left.
- This behavior is symmetric in time.

Example: Random Walk on a Graph

- A **graph** G is a countable collection of states, usually called **vertices**, some of which are joined by **edges**.



- Thus a graph is a partially drawn Markov chain diagram.
- There is a natural way to complete the diagram which gives rise to the random walk on G.

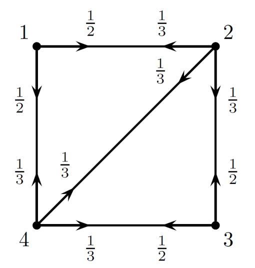
Example: Random Walk on a Graph (Cont'd)

- The **valency** v_i of vertex i is the number of edges at i.

  We assume that every vertex has finite valency.

  The random walk on G picks edges with equal probability. Thus, the transition probabilities are given by

$$p_{ij} = \begin{cases} \frac{1}{v_i}, & \text{if } (i,j) \text{ is an edge,} \\ 0, & \text{otherwise.} \end{cases}$$



  We assume G is connected, so that P is irreducible.

  We may show that P is in detailed balance with $v = (v_i : i \in G)$.

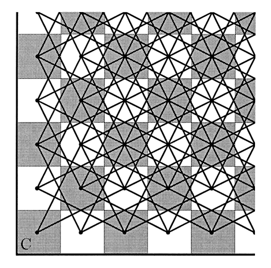
  Suppose the total valency $\sigma = \sum_{i \in G} v_i$ is finite.

  Then $\pi = \frac{v}{\sigma}$ is invariant and P is reversible.

Example: Random Chessboard Knight

- A random knight makes each permissible move with equal probability. If it starts in a corner, how long on average will it take to return?

- This is an example of a random walk on a graph.
- The vertices are the squares of the chessboard.
- The edges are the moves that the knight can take.
- The diagram shows a part of the graph.



- We know by a previous theorem and the preceding example that

$$\mathbb{E}_c(T_c) = \frac{1}{\pi_c} = \frac{1}{v_c/\sigma} = \frac{\sum_i v_i}{v_c}.$$

Example: Random Chessboard Knight (Cont'd)

- We have
$$\mathbb{E}_c(T_c) = \frac{\sum_i v_i}{v_c}.$$
- So all we have to do is identify valencies.
  - The four corner squares have valency 2.
  - The eight squares adjacent to the corners have valency 3.
  - There are 20 squares of valency 4
  - There are 16 squares of valency 6
  - The 16 central squares have valency 8.
- Hence
$$\mathbb{E}_c(T_c) = \frac{8+24+80+96+128}{2} = 168.$$

Subsection 11

Ergodic Theorem

Strong Law of Large Numbers

Theorem (Strong Law of Large Numbers)

Let $Y_1, Y_2, \ldots$ be a sequence of independent, identically distributed, non-negative random variables with $\mathbb{E}(Y_1) = \mu$. Then

$$\mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

- A proof for $\mu < \infty$ is found in standard probability texts.
  The case where $\mu = \infty$ is a simple deduction.
  Fix $N < \infty$. Set $Y_n^{(N)} = Y_n \wedge N$. Then

$$\begin{aligned}\frac{Y_1+\cdots+Y_n}{n} &\geq \frac{Y_1^{(N)}+\cdots+Y_n^{(N)}}{n}\\ &\to \mathbb{E}(Y_1 \wedge N), \text{ as } n \to \infty,\\ &\ \text{with probability one.}\end{aligned}$$

  As $N \to \infty$ we have $\mathbb{E}(Y_1 \wedge N) \nearrow \mu$ by monotone convergence.
  So, with probability 1, $\frac{Y_1+\cdots+Y_n}{n} \to \infty$ as $n \to \infty$.

Number of Visits Before Time n

- We denote by $V_i(n)$ the **number of visits to** i **before** n:

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k=i\}}.$$

- Then $\frac{V_i(n)}{n}$ is the proportion of time before n spent in state i.

The Ergodic Theorem

Theorem (Ergodic Theorem)

Let P be irreducible and let λ be any distribution. If $(X_n)_{n\geq 0}$ is Markov(λ, P), then

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1,$$

where $m_i = \mathbb{E}_i(T_i)$ is the expected return time to state i. Moreover, in the positive recurrent case, for any bounded function $f : I \to \mathbb{R}$, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \overline{f} \text{ as } n \to \infty\right) = 1,$$

where $\overline{f} = \sum_{i \in I} \pi_i f_i$ and where $(\pi_i : i \in I)$ is the unique invariant distribution.

Proof of the Ergodic Theorem

- If P is transient, then, with probability 1, the total number V_i of visits to i is finite. So

$$\frac{V_i(n)}{n} \leq \frac{V_i}{n} \to 0 = \frac{1}{m_i}.$$

Suppose then that P is recurrent and fix a state i.

For $T = T_i$ we have:

  - $P(T < \infty) = 1$, by a previous theorem;
  - $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \ldots, X_T$, by the Strong Markov Property.

The long run proportion of time spent in i is the same for $(X_{T+n})_{n \geq 0}$ and $(X_n)_{n \geq 0}$.

So it suffices to consider the case $\lambda = \delta_i$.

Proof of the Ergodic Theorem (Cont'd)

- Write $S_i^{(r)}$ for the length of the r-th excursion to i.

  By a previous lemma, the non-negative random variables $S_i^{(1)}, S_i^{(2)}, \ldots$ are independent and identically distributed with $\mathbb{E}_i(S_i^{(r)}) = m_i$.

  $S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)}$ is the time of the last visit to i before n.

  So we have
  $$S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)} \leq n - 1.$$

  $S_i^{(1)} + \cdots + S_i^{(V_i(n))}$ is the time of the first visit to i after $n-1$.

  So we have
  $$S_i^{(1)} + \cdots + S_i^{(V_i(n))} \geq n.$$

  These give
  $$\frac{S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{S_i^{(1)} + \cdots + S_i^{(V_i(n))}}{V_i(n)}.$$

Proof of the Ergodic Theorem (Cont'd)

- We got

$$\frac{S_i^{(1)} + \cdots + S_i^{(V_i(n)-1)}}{V_i(n)} \leq \frac{n}{V_i(n)} \leq \frac{S_i^{(1)} + \cdots + S_i^{(V_i(n))}}{V_i(n)}.$$

By the strong law of large numbers

$$\mathbb{P}\left(\frac{S_i^{(1)} + \cdots + S_i^{(n)}}{n} \to m_i \text{ as } n \to \infty\right) = 1.$$

Since P is recurrent,

$$\mathbb{P}\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1.$$

This implies

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1.$$

Proof of the Ergodic Theorem (Conclusion)

- Assume now that $(X_n)_{n\geq 0}$ has an invariant distribution $(\pi_i : i \in I)$.
  Let $f : I \to \mathbb{R}$ be a bounded function.
  Assume without loss of generality that $|f| \leq 1$.
  For any $J \subseteq I$, we have

$$\begin{array}{rcl} |\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}| & = & |\sum_{i\in I}(\frac{V_i(n)}{n} - \pi_i)f_i| \\ & \leq & \sum_{i\in J}|\frac{V_i(n)}{n} - \pi_i| + \sum_{i\notin J}|\frac{V_i(n)}{n} - \pi_i| \\ & \leq & \sum_{i\in J}|\frac{V_i(n)}{n} - \pi_i| + \sum_{i\notin J}(\frac{V_i(n)}{n} + \pi_i) \\ & \leq & 2\sum_{i\in J}|\frac{V_i(n)}{n} - \pi_i| + 2\sum_{i\notin J}\pi_i. \end{array}$$

  We proved above that $\mathbb{P}\left(\frac{V_i(n)}{n} \to \pi_i \text{ as } n \to \infty \text{ for all } i\right) = 1$.
  Given $\varepsilon > 0$, choose J finite so that $\sum_{i\notin J}\pi_i < \frac{\varepsilon}{4}$.
  Then choose $N = N(\omega)$ so that, for $n \geq N(\omega)$, $\sum_{i\in J}|\frac{V_i(n)}{n} - \pi_i| < \frac{\varepsilon}{4}$.
  Then, for $n \geq N(\omega)$, we have $|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}| < \varepsilon$.
  This establishes the desired convergence.

Estimating Transition Probabilities

- Sometimes we need to estimate an unknown transition matrix P on the basis of observations of the corresponding Markov chain.
- Consider the case where we have $N+1$ observations $(X_n)_{0 \leq n \leq N}$.
- The log-likelihood function is given by

$$\ell(P) = \log\left(\lambda_{X_0} p_{X_0 X_1} \cdots p_{X_{N-1} X_N}\right) = \sum_{i,j \in I} N_{ij} \log p_{ij}$$

up to a constant independent of P, where N_{ij} is the number of transitions from i to j.

Estimating Transition Probabilities (Cont'd)

- A standard statistical procedure is to find the **maximum likelihood estimate** $\widehat{P}$, which is the choice of P maximizing $\ell(P)$.
- P must satisfy the linear constraint $\sum_j p_{ij} = 1$, for each i.
- So we first try to maximize
$$\ell(P) + \sum_{i,j \in I} \mu_i p_{ij}$$
and then choose $(\mu_i : i \in I)$ to fit the constraints.
- This is the method of Lagrange multipliers.
- Thus we find
$$\widehat{p}_{ij} = \frac{\sum_{n=0}^{N-1} 1_{\{X_n = i, X_{n+1} = j\}}}{\sum_{n=0}^{N-1} 1_{\{X_n = i\}}},$$
which is the proportion of jumps from i which go to j.

Consistency of the Estimate

- We now consider the **consistency** of this sort of estimate, i.e., whether $\widehat{p}_{ij} \to p_{ij}$, with probability 1, as $N \to \infty$.
- This is clearly false when i is transient.
- So we shall slightly modify our approach.
- Note that to find $\widehat{p}_{ij}$ we simply have to maximize $\sum_{j \in I} N_{ij} \log p_{ij}$ subject to $\sum_j p_{ij} = 1$, the other terms and constraints being irrelevant.
- Suppose then that instead of $N + 1$ observations we make enough observations to ensure the chain leaves state i a total of N times.
- In the transient case this may involve restarting the chain several times.
- Denote again by N_{ij} the number of transitions from i to j.

Consistency of the Estimate (Cont'd)

- To maximize the likelihood for $(p_{ij} : j \in I)$ we still maximize

$$\sum_{j \in I} N_{ij} \log p_{ij}$$

  subject to $\sum_j p_{ij} = 1$.
- This leads to the maximum likelihood estimate $\widehat{p}_{ij} = \frac{N_{ij}}{N}$.
- But $N_{ij} = Y_1 + \cdots + Y_N$, where $Y_n = 1$ if the n-th transition from i is to j, and $Y_n = 0$ otherwise.
- By the strong Markov property $Y_1, \ldots, Y_N$ are independent and identically distributed random variables with mean p_{ij}.
- So, by the strong law of large numbers

$$\mathbb{P}(\widehat{p}_{ij} \to p_{ij} \text{ as } N \to \infty) = 1.$$

- This shows that $\widehat{p}_{ij}$ is consistent.