Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Laboratorio de Ciberseguridad

Centro de Investigación en Computación(CIC)
Instituto Politécnico Nacional - México.

Probability, Random Processes and Inference

Dr. Ponciano Jorge Escamilla Ambrosio
pescamilla@cic.ipn.mx
http://www.cic.ipn.mx/~pescamilla/

Course Content

2. Introduction to Random Processes
   - 2.1. Markov Chains
     - 2.1.1. Discrete Time Markov Chains
     - 2.1.2. Classification of States
     - 2.1.3. Steady State Behavior
     - 2.1.4. Absorption Probabilities and Expected Time to Absorption
     - 2.1.5. Continuous Time Markov Chains
     - 2.1.6. Ergodic Theorem for Discrete Markov Chains
     - 2.1.7. Markov Chain Montecarlo Method
     - 2.1.8. Queuing Theory

Stochastic (Random) Processes

- A random or stochastic process is a mathematical model for a phenomenon that evolves in time in an unpredictable manner from the viewpoint of the observer.
- It may be unpredictable because of such effects as interference or noise in a communication link or storage medium, or it may be an information-bearing signal, deterministic from the viewpoint of an observer at the transmitter but random to an observer at the receiver.

Stochastic (Random) Processes

- A stochastic (or random) process is a mathematical model of a probabilistic experiment that evolves in time and generates a sequence of numerical values.
- A stochastic process can be used to model:
  - The sequence of daily prices of a stock;
  - The sequence of scores in a football game;
  - The sequence of failure times of a machine;
  - The sequence of hourly traffic loads at a node of a communication network;
  - The sequence of radar measurements of the position of an airplane.

Stochastic (Random) Processes

- Each numerical value in the sequence is modelled by a random variable.
- A stochastic process is simply a (finite or infinite) sequence of random variables.
- We are still dealing with a single basic experiment that involves outcomes governed by a probability law, and random variables that inherit their probabilistic properties from that law.

Stochastic (Random) Processes

- Stochastic processes involve some changes with respect to earlier models:
  - Tend to focus on the dependencies in the sequence of values generated by the process.
    - How do future prices of a stock depend on past values?
  - Are often interested in long-term averages involving the entire sequence of generated values.
    - What is the fraction of time that a machine is idle?
  - Wish to characterize the likelihood or frequency of certain boundary events.
    - What is the probability that within a given hour all circuits of some telephone system become simultaneously busy?

Stochastic (Random) Processes

- There is a wide variety of stochastic processes.
- Two major categories are of concern of this course:
  - **Arrival-Type Processes.** The occurrences have the character of an "arrival", such as message receptions at a receiver, job completions in a manufacturing cell, etc. These are models in which the interarrival times (the times between successive arrivals) are independent random variables.
    - **Bernoulli process**. Arrivals occur in discrete times and the interarrival times are geometrically distributed.
    - **Poisson process.** Arrivals occur in continuous time and the interarrivals times are exponentially distributed.

Stochastic (Random) Processes

- Two major categories are of concern of this course:
  - Markov Processes. Involve experiments that evolve in time and which the future evolution exhibits a probabilistic dependence on the past. As an example, the future daily prices of a stock are typically dependent on past prices.
    - In a Markov process, it is assumed a very special type of dependence: the next value depends on past values only through the current value.

Discrete-Time Markov Chains

- Markov chains were first introduced in 1906 by Andrey Markov (of Markov's inequality), with the goal of showing that the law of large numbers can apply to random variables that are not independent.
- Markov began the study of an important new type of chance process. In this process, the outcome of a given experiment can affect the outcome of the next experiment. This type of process is called a Markov chain.

Discrete-Time Markov Chains

- Since their invention, Markov chains have become extremely important in a huge number of fields such as biology, game theory, finance, machine learning, and statistical physics.
- They are also very widely used for simulations of complex distributions, via algorithms known as Markov chain Monte Carlo (MCMC).

Discrete-Time Markov Chains

- Markov chains "live" in both space and time: the set of possible values of the X_n is called the state space, and the index n represents the evolution of the process over time.
- The state space of a Markov chain can be either discrete or continuous, and time can also be either discrete or continuous.
  - in the continuous-time setting, we would imagine a process X_t defined for all real $t \geq 0$.

Discrete-Time Markov Chains

- Consider first discrete-time Markov chains, in which the state changes at certain discrete time instants, indexed by an integer variable *n*.
- At each time step *n*, the state of the chain is denoted by X_n and belongs to a finite set *S* of possible states, called **state space**.
- Specifically, we will assume that $S = \{1, 2, \ldots, m\}$, for some positive integer *m*.

Discrete-Time Markov Chains

- Formally, a sequence of random variables X_0, X_1, X_2,… taking values in the state space $S = \{1, 2,\ldots, m\}$ is called a **Markov chain** if for all $n \geq 0$, the **Markov property** is satisfied:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$$
$$= p_{ij}$$

  - The quantity $p_{ij} = P(X_{n+1} = j \mid X_n = i)$ is called the **transition probability** from state i to state j, with $i, j \in S$, for all times n, and all possible sequences $i_0,\ldots, i_{n-1}$ of earlier states.

Discrete-Time Markov Chains

- Whenever the state happens to be i, there is probability p_{ij} that the next state is equal to j.
- The key assumption underlying the Markov chains is that the transition probabilities p_{ij} apply whenever state i is visited, no matter what happened in the past, and no matter how state i was reached.
- **The probability law of the next state** X_{n+1} **depends on the past only through the value of the present state** X_n.

Discrete-Time Markov Chains

- In other words, given the entire past history X_0, X_1, X_2,…, X_n, only the most recent term, X_n, matters for predicting X_{n+1}.
- If we think of time *n* as the present, times before *n* as the past, and times after *n* as the future, the Markov property says that given the present, the past and future are conditionally independent.
- The transition probabilities p_{ij} must be nonnegative, and sum to one:

$$\sum_{j=1}^{m} p_{ij} = 1, \qquad \text{for all } i$$

Discrete-Time Markov Chains

Specification of Markov Models

- A Markov chain model is specified by identifying:
  - (a) the set of states $\mathcal{S} = \{1, \ldots, m\}$,
  - (b) the set of possible transitions, namely, those pairs (i, j) for which $p_{ij} > 0$, and,
  - (c) the numerical values of those p_{ij} that are positive.
- The Markov chain specified by this model is a sequence of random variables $X_0, X_1, X_2, \ldots$, that take values in $\mathcal{S}$, and which satisfy

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = p_{ij},$$

for all times n, all states $i, j \in \mathcal{S}$, and all possible sequences $i_0, \ldots, i_{n-1}$ of earlier states.

Discrete-Time Markov Chains

- All of the elements of a Markov chain model can be encoded in a **transition probability matrix P**, whose (i, j) entry is the probability of going from state i to state j in one step of the chain.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

  - Note that P is a nonnegative matrix in which each row sums to 1.

Discrete-Time Markov Chains

- It is also helpful to lay out the model in the so-called **transition probability graph**, whose nodes are the states and whose arcs are the possible transitions.
- By recording the numerical values of p_{ij} near the corresponding arcs, one can visualize the entire model in a way that can make some of its major properties readily apparent.

Discrete-Time Markov Chains

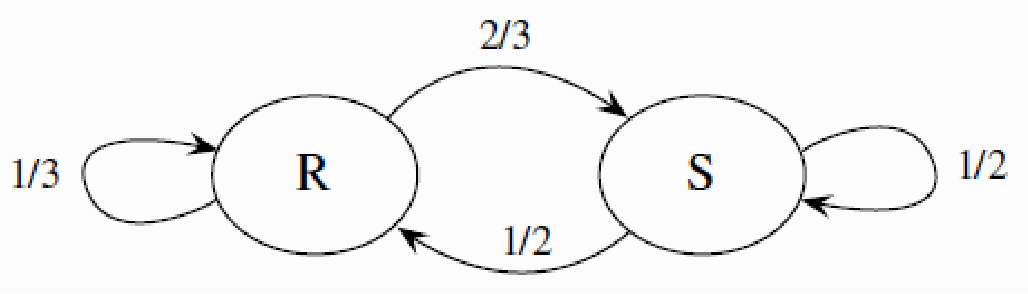
- Example 1. Rainy-sunny Markov chain. Suppose that on any given day, the weather can either be rainy or sunny. If today is rainy, then tomorrow will be rainy with probability 1/3 and sunny with probability 2/3. If today is sunny, then tomorrow will be rainy with probability 1/2 and sunny with probability 1/2. Letting X_n be the weather on day *n*, $X_0, X_1, X_2, \ldots$ is a Markov chain on the state space {R, S}, where R stands for rainy and S for sunny. We know that the Markov property is satisfied because, from the description of the process, only today's weather matters for predicting tomorrow's.

Discrete-Time Markov Chains

- Transition probability matrix:

$$
\begin{array}{cc}
 & \begin{array}{cc} R & S \end{array} \\
\begin{array}{c} R \\ S \end{array} & \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}
\end{array}
$$

- Transition probability graph:



Discrete-Time Markov Chains

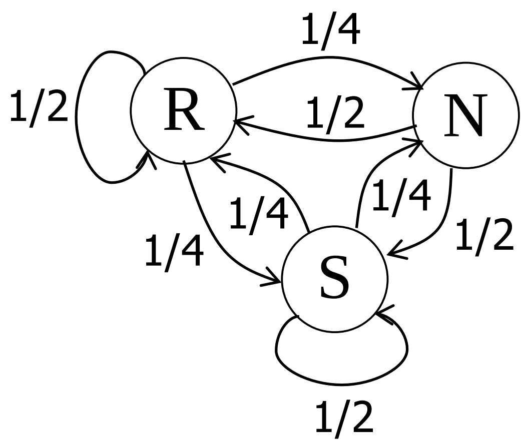
- Example 2. According to Kemeny, Snell, and Thompson, the Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. With this information, form a Markov chain model.

Discrete-Time Markov Chains

- Transition probability matrix:

$$\mathbf{P} = \begin{array}{c} \\ \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

- Transition probability graph:



Discrete-Time Markov Chains

- Example 3. Spiders and Fly. A fly moves along a straight line in unit increments. At each time period, it moves one unit to the left with probability 0.3, one unit to the right with probability 0.3, and stays in place with probability 0.4, independent of the past history of movements. Two spiders are lurking at positions 1 and m; if the fly lands there, it is captured by a spider, and the process terminates. Construct a Markov chain model, assuming that the fly starts in a position between 1 and m.

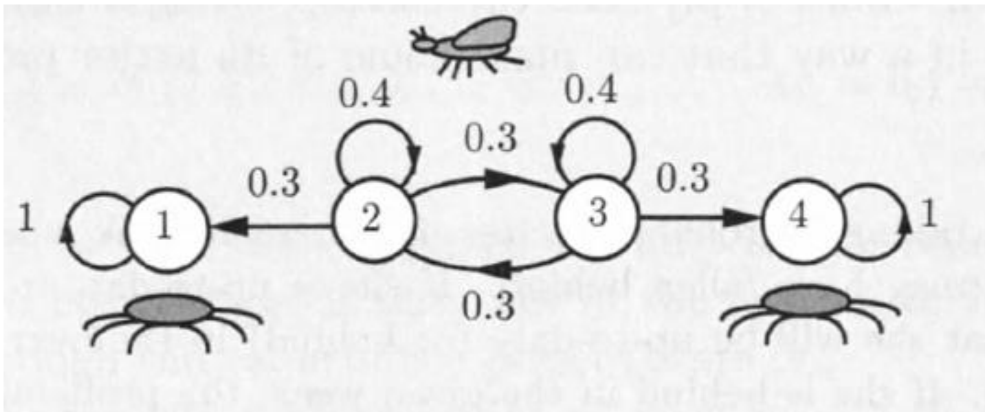
Discrete-Time Markov Chains

Let us introduce states $1, 2, \ldots, m$, and identify them with the corresponding positions of the fly. The nonzero transition probabilities are

$$p_{11} = 1, \qquad p_{mm} = 1,$$

$$p_{ij} = \begin{cases} 0.3, & \text{if } j = i - 1 \text{ or } j = i + 1, \\ 0.4, & \text{if } j = i, \end{cases} \qquad \text{for } i = 2, \ldots, m - 1.$$

The transition probability graph and matrix are shown in Fig. 7.2.



	1	2	3	4
1	1.0	0	0	0
2	0.3	0.4	0.3	0
3	0	0.3	0.4	0.3
4	0	0	0	1.0

p_{ij}

Figure 7.2: The transition probability graph and the transition probability matrix in Example 7.2, for the case where $m = 4$.

Discrete-Time Markov Chains

- The probability of a path. Given a Markov chain model, can compute the probability of any particular sequence of future states.
  - This is analogous to the use of the multiplication rule in sequential (tree) probability models.

Discrete-Time Markov Chains

- In particular: Let be $P(n) = \{p_{ij}(n)\}$ be the matrix of n-step transition probabilities, where:

$$p_{ij}(n) = P[X_{n+k} = j \mid X_k = i] \qquad n \geq 0, i, j \geq 0.$$

  - Where $p_{ij}(n)$ is the probability that the state after n time periods will be j, given that the current state is i.
- Note that $P[X_{n+k} = j \mid X_k = i] = P[X_n = j \mid X_0 = i]$ for all $n \geq 0$ and $k \geq 0$, since the transition probabilities do not depend on time.

Discrete-Time Markov Chains

- First, consider the two-step transition probabilities. The probability of going from state *i* at *t* = 0 passing through state *k* at *t* = 1, and ending at state *j* at *t* = 2 is:

$$
\begin{aligned}
P[X_2 = j, X_1 = k \mid X_0 = i] &= \frac{P[X_2 = j, X_1 = k, X_0 = i]}{P[X_0 = i]} \\
&= \frac{P[X_2 = j \mid X_1 = k] P[X_1 = k \mid X_0 = i] P[X_0 = i]}{P[X_0 = i]} \\
&= P[X_2 = j \mid X_1 = k] P[X_1 = k \mid X_0 = i] \\
&= p_{ik}(1) p_{kj}(1).
\end{aligned}
$$

Discrete-Time Markov Chains

- Note that $p_{ik}(1)$ and $p_{kj}(1)$ are components of P, the one-step transition probability matrix. We obtain $p_{ij}(2)$, the probability of going from i at t = 0 to j at t = 2, by summing over all possible intermediate states k:

$$p_{ij}(2) = \sum_k p_{ik}(1) p_{kj}(1) \qquad \text{for all } i, j.$$

  - This is, the *ij* entry of P(2) is obtained by multiplying the *i*th row of P(1) by the *j*th column of P(1). In other words, P(2) is obtained by multiplying the one-step transition probability matrices:

$$P(2) = P(1)P(1) = P^2.$$

Discrete-Time Markov Chains

- Now consider the probability of going from state i at $t = 0$, passing through state k at $t = m$, and ending at state j at time $t = m + n$. Following the same procedure as above we obtain the **Chapman–Kolmogorov equations**:

$$p_{ij}(m+n) = \sum_k p_{ik}(m) p_{kj}(n) \quad \text{for all } n, m \geq 0 \text{ all } i, j.$$

Discrete-Time Markov Chains

- Therefore the matrix of $n + m$ step transition probabilities $P(n + m) = \{p_{ij}(n + m)\}$ is obtained by the following matrix multiplication:

$$P(n + m) = P(n)P(m)$$

- By induction, this implies that:

$$P(n) = P^n$$

Discrete-Time Markov Chains

Chapman-Kolmogorov Equation for the n-Step Transition Probabilities

The n-step transition probabilities can be generated by the recursive formula

$$r_{ij}(n) = \sum_{k=1}^{m} r_{ik}(n-1)p_{kj}, \qquad \text{for } n > 1, \quad \text{and all } i,\ j,$$

starting with

$$r_{ij}(1) = p_{ij}.$$

Discrete-Time Markov Chains

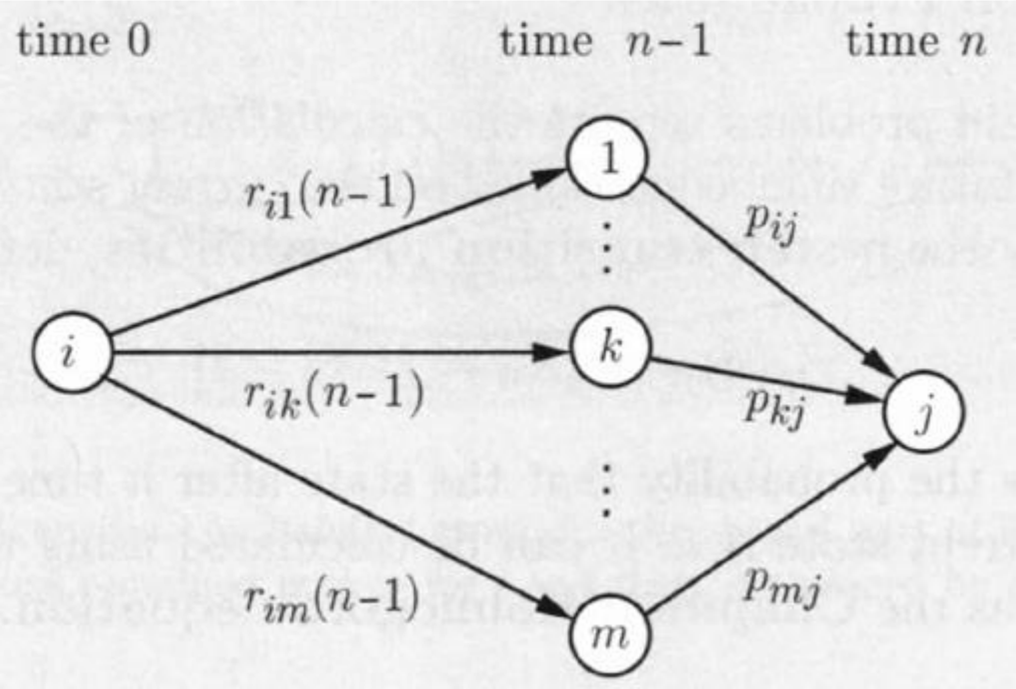
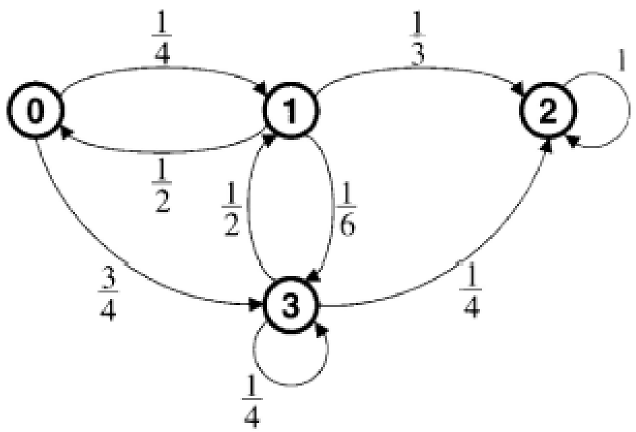


Figure 7.5: Derivation of the Chapman-Kolmogorov equation. The probability of being at state j at time n is the sum of the probabilities $r_{ik}(n-1)p_{kj}$ of the different ways of reaching j.

Discrete-Time Markov Chains

Example



$$P = \begin{bmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{bmatrix}$$

Consider the probability of going from state 0 to state 3 in exactly 3 steps. From the graph, all possible paths are

$$0-1-0-3, 0-1-3-3, 0-3-1-3, \text{ and } 0-3-3-3$$

Probability of success for each path is: 3/32, 1/96, 1/16 and 3/64 respectively. Summing up the probabilities we find the total probability is 41/192.

Discrete-Time Markov Chains

Example

Alternatively, we can compute

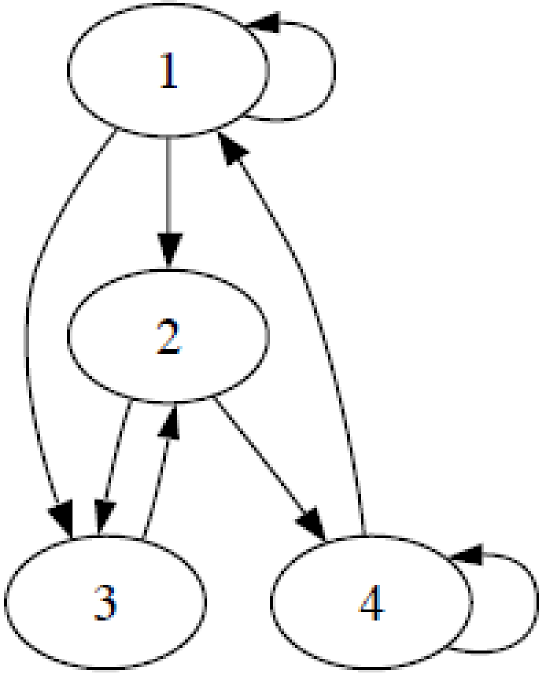
$$P^3 = \begin{bmatrix} 3/16 & 7/48 & 29/64 & 41/192 \\ 5/48 & 5/24 & 79/144 & 5/36 \\ 0 & 0 & 1 & 0 \\ 1/16 & 13/96 & 107/192 & 47/192 \end{bmatrix}$$

The entry $P^3_{0,3} = 41/192$ gives the correct answer.

Discrete-Time Markov Chains

- Example 4. Transition matrix of 4-state Markov chain. Consider the 4-state Markov chain depicted in the Figure. When no probabilities are written over the arrows, as in this case, it means all arrows originating from a given state are equally likely. For example, there are 3 arrows originating from state 1, so the transitions $1 \to 3$, $1 \to 2$, and $1 \to 1$ all have probability 1/3. (a) what is the transition matrix? (b) what is the probability that the chain is in state 3 after 5 steps, starting at state 1?

Discrete-Time Markov Chains

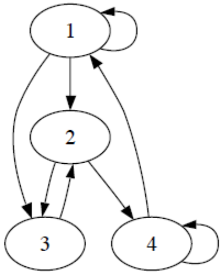


Discrete-Time Markov Chains

- Transition probability matrix:

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

$$P^5 = \begin{pmatrix} 853/3888 & 509/1944 & 52/243 & 395/1296 \\ 173/864 & 85/432 & 31/108 & 91/288 \\ 37/144 & 29/72 & 1/9 & 11/48 \\ 499/2592 & 395/1296 & 71/324 & 245/864 \end{pmatrix}$$



so $p_{13}^{(5)} = 52/243.$

Discrete-Time Markov Chains

- We now consider the long-term behavior of a Markov chain when it starts in a state chosen by a probability distribution on the set of states, which we will call a probability vector.
- A probability vector with r components is a row vector whose entries are non-negative and sum to 1.
- If $\mathbf{u}$ is a probability vector which represents the initial state of a Markov chain, then we think of the ith component of $\mathbf{u}$ as representing the probability that the chain starts in state s_i.

Discrete-Time Markov Chains

- Let P be the transition matrix of a Markov chain, and let **u** be the probability vector which represents the starting distribution. Then the probability that the chain is in state s_i after n steps is the ith entry in the vector:

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$$

  - We note that if we want to examine the behavior of the chain under the assumption that it starts in a certain state s_i, we simply choose **u** to be the probability vector with ith entry equal to 1 and all other entries equal to 0.

Discrete-Time Markov Chains

- Example 5. In the Land of Oz example (Example 2) let the initial probability vector **u** equal (1/3, 1/3, 1/3), meaning that the chain has equal probability of starting in each of the three states. Calculate the distribution of the states after three days.

$$
\mathbf{P} = \begin{array}{c} \\ \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array}
\begin{array}{c} \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}
$$

Discrete-Time Markov Chains

- Example 5. In the Land of Oz example (Example 2) let the initial probability vector **u** equal (1/3, 1/3, 1/3), meaning that the chain has equal probability of starting in each of the three states. Calculate the distribution of the states after three days.

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{ccc} \mathrm{R} & \mathrm{N} & \mathrm{S} \end{array} \\ \begin{array}{c} \mathrm{R} \\ \mathrm{N} \\ \mathrm{S} \end{array} & \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}$$

$$\begin{aligned} \mathbf{u}^{(3)} = \mathbf{uP}^3 &= \begin{pmatrix} 1/3, & 1/3, & 1/3 \end{pmatrix} \begin{pmatrix} .406 & .203 & .391 \\ .406 & .188 & .406 \\ .391 & .203 & .406 \end{pmatrix} \\ &= \begin{pmatrix} .401, & .198, & .401 \end{pmatrix} . \end{aligned}$$

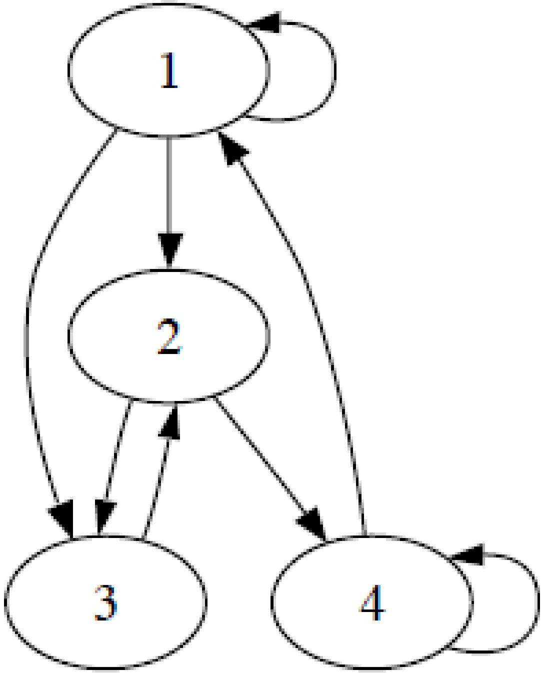
Discrete-Time Markov Chains

- Example 6. Consider the 4-state Markov chain in Example 4. Suppose the initial conditions are **t** =

(1/4, 1/4, 1/4, 1/4), meaning that the chain has equal probability of starting in each of the four states. Let X_n be the position of the chain at time n. Then the marginal distribution of X_5 is:

Discrete-Time Markov Chains

- Example 6.



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

Discrete-Time Markov Chains

- Example 6.

$$tQ = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

$$= \begin{pmatrix} 5/24 & 1/3 & 5/24 & 1/4 \end{pmatrix}.$$

The marginal distribution of X_5 is

$$tQ^5 = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 853/3888 & 509/1944 & 52/243 & 395/1296 \\ 173/864 & 85/432 & 31/108 & 91/288 \\ 37/144 & 29/72 & 1/9 & 11/48 \\ 499/2592 & 395/1296 & 71/324 & 245/864 \end{pmatrix}$$

$$= \begin{pmatrix} 3379/15552 & 2267/7776 & 101/486 & 1469/5184 \end{pmatrix}.$$

Classification of States

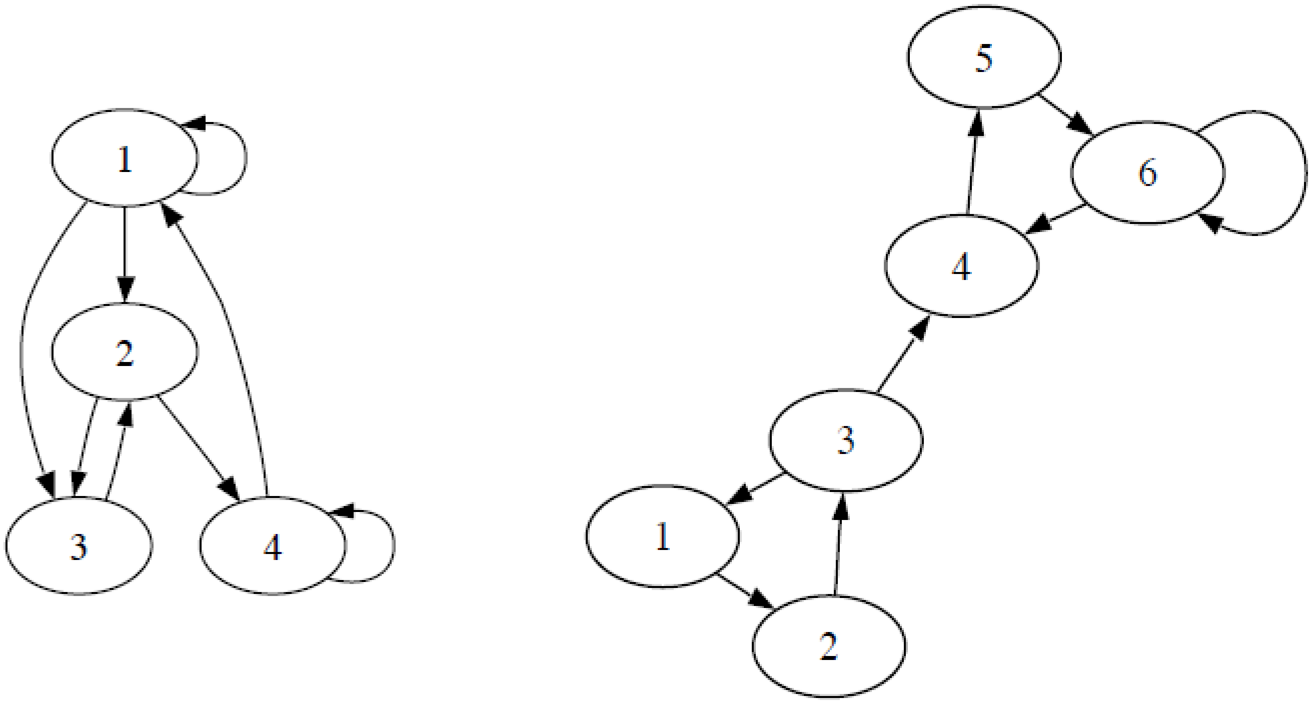
- The states of a Markov chain can be classified as **recurrent** or **transient**, depending on whether they are visited over and over again in the long run or are eventually abandoned.
- States can also be **classified according to their period**, which is a positive integer summarizing the amount of time that can elapse between successive visits to a state.

Classification of States

- Recurrent and transient states.
  - State *i* of a Markov chain is **recurrent** if starting from *i*, the probability is 1 that the chain will eventually return to *i*.
  - Otherwise, the state is **transient**, which means that if the chain starts from *i*, there is a positive probability of never returning to *i*.
- As long as there is a positive probability of leaving *i* forever, the chain eventually will leave *i* forever.

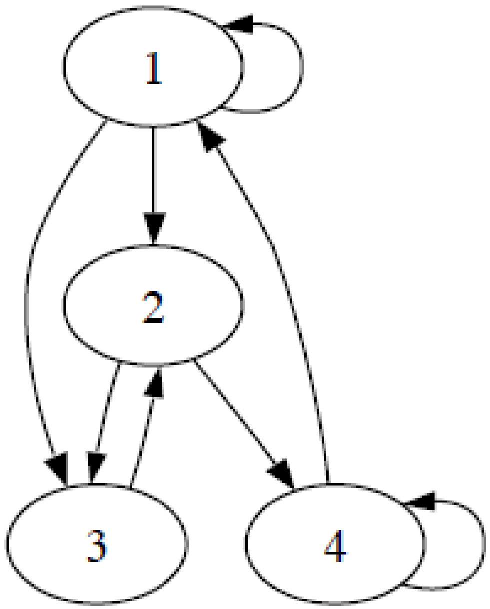
Classification of States

- Example 7. In the Markov chains shown below, are the states recurrent or transient?



Classification of States

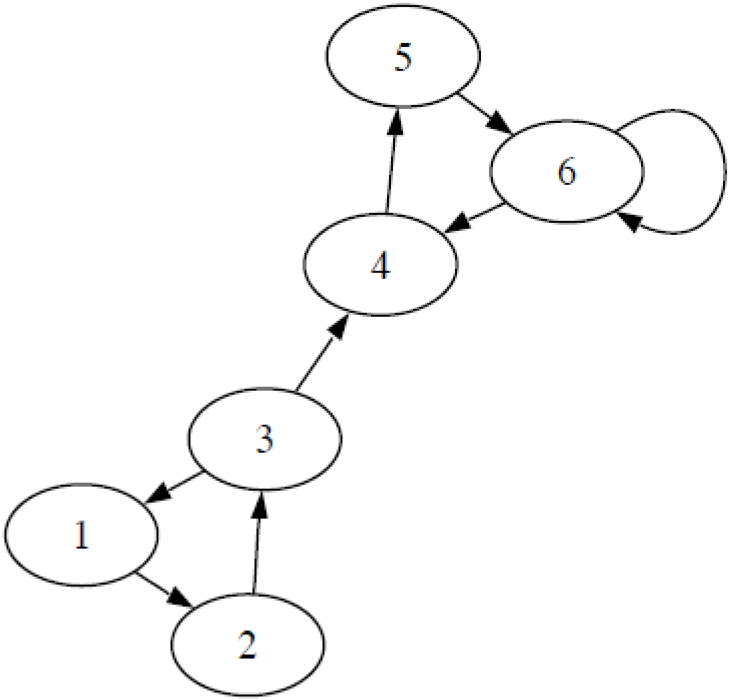
- Example 7. In the Markov chains shown below, are the states recurrent or transient?



A particle moving around between states will continue to spend time in all 4 states in the long run, since it is possible to get from any state to any other state.

Classification of States

- Example 7. In the Markov chains shown below, are the states recurrent or transient?



Let the particle start at state 1. For a while, the chain may linger in the triangle formed by states 1, 2, and 3, but eventually it will reach state 4, and from there it can never return to states 1, 2, or 3. It will then wander around between states 4, 5, and 6 forever. States 1, 2, and 3 are transient and states 4, 5, and 6 are recurrent.

Classification of States

- Although the definition of a transient state only requires that there be a positive probability of never returning to the state, we can say something stronger:
  - As long as there is a positive probability of leaving *i* forever, the chain eventually will leave *i* forever.
  - In the long run, anything that can happen, will happen (with a finite state space).

Classification of States

- A state j is **accessible** from state i if for some n, the n-step transition probability $p_{ij}(n)$ is positive, i.e., if there is positive probability of reaching j, starting from i, after some number of time periods.
- Let $A(i)$ be the set of states that are accessible from i.
  - i is recurrent if for every j that is accessible from i, i also is accessible from j; that is, for all j that belong to $A(i)$ we have that i belongs to $A(j)$.

Classification of States

- If i is a recurrent state, the set of states A(i) that are accessible from i form a **recurrent class** (or simply **class**), meaning that states in A(i) are all accessible from each other, and no state outside A(i) is accessible from them.
- Mathematically, for a recurrent state i, we have A(i) = A(j) for all j that belong to A(i), as can be seen from the definition of recurrence.

Classification of States

- Example 8.

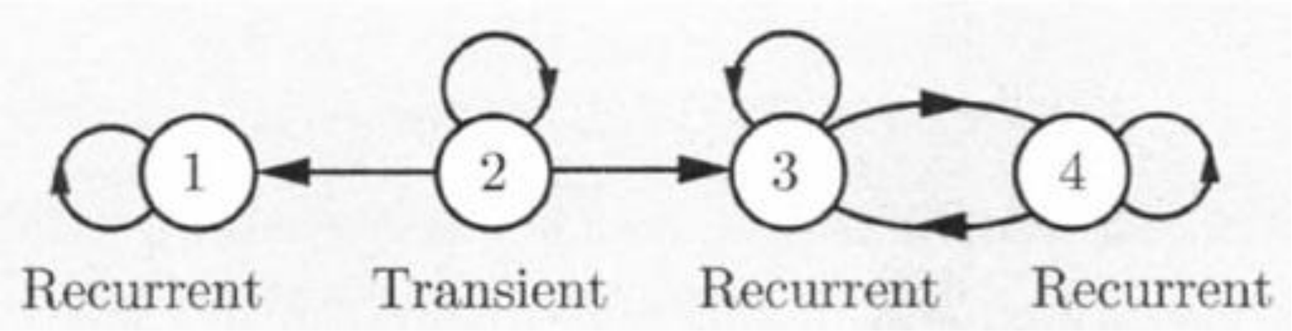


Figure 7.8: Classification of states given the transition probability graph. Starting from state 1, the only accessible state is itself, and so 1 is a recurrent state. States 1, 3, and 4 are accessible from 2, but 2 is not accessible from any of them, so state 2 is transient. States 3 and 4 are accessible from each other, and they are both recurrent.

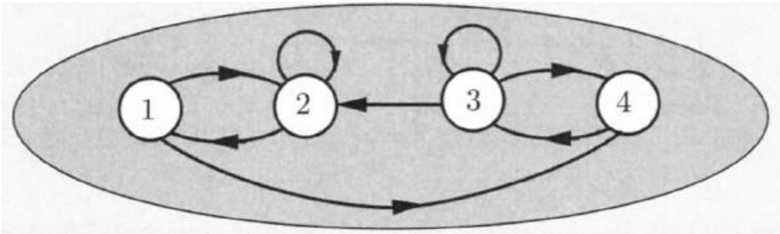
Classification of States

Markov Chain Decomposition

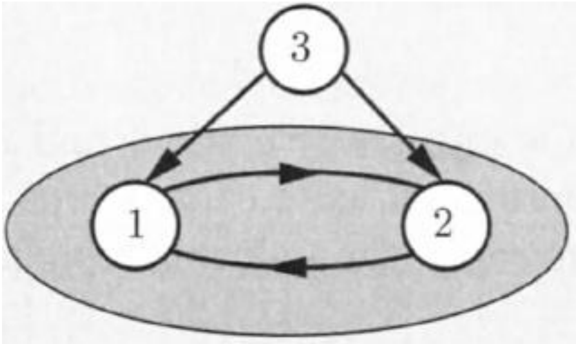
- A Markov chain can be decomposed into one or more recurrent classes, plus possibly some transient states.
- A recurrent state is accessible from all states in its class, but is not accessible from recurrent states in other classes.
- A transient state is not accessible from any recurrent state.
- At least one, possibly more, recurrent states are accessible from a given transient state.

Classification of States

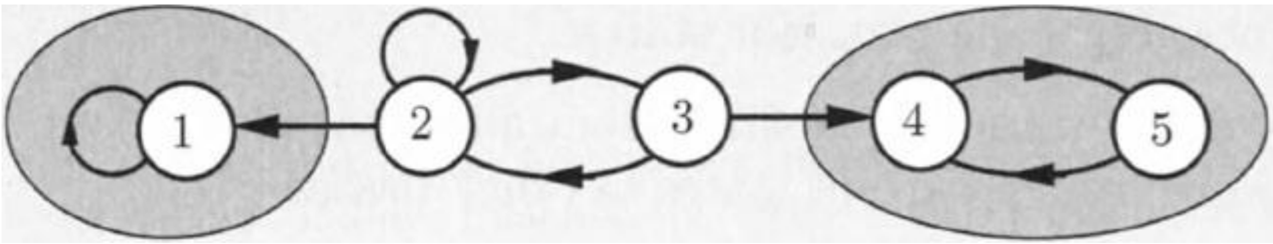
- Examples of Markov chain decompositions:



Single class of recurrent states



Single class of recurrent states (1 and 2) and one transient state (3)



Two classes of recurrent states (class of state 1 and class of states 4 and 5) and two transient states (2 and 3)

Classification of States

- From Markov chain decomposition:
  - (a) once the state enters (or starts in) a class of recurrent states, it stays within that class; since all states in the class are accessible from each other, all states in the class will be visited an infinite number of times
  - (b) if the initial state is transient, then the state trajectory contains an initial portion consisting of transient states and a final portion consisting of recurrent states from the same class.

Classification of States

- For the purpose of understanding **long-term behaviour** of Markov chains, it is important to **analyse chains that consist of a single recurrent class**.
- For the purpose of understanding **short-term behaviour**, it is also important to **analyse the mechanism by which any particular class of recurrent states is entered starting from a given transient state**.

Classification of States

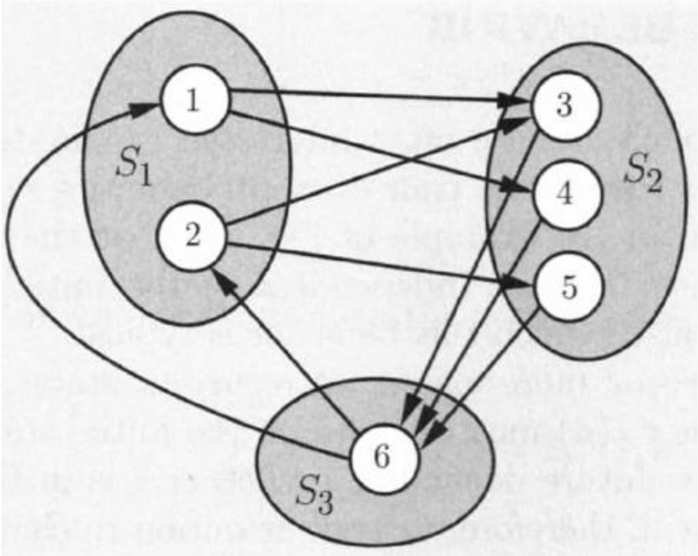
- Periodicity. A recurrent class is said to be **periodic** if its states can be grouped in $d > 1$ disjoint subsets $S_1, \ldots, S_d$ so that all transitions from one subset lead to the next subset:

$$\text{if} \quad i \in S_k \text{ and } p_{ij} > 0, \quad \text{then} \begin{cases} j \in S_{k+1}, & \text{if } k = 1, \ldots, d-1, \\ j \in S_1, & \text{if } k = d. \end{cases}$$

- A recurrent class that is not periodic, is said to be **aperiodic**.

Classification of States

- In a periodic recurrent class, we move through the sequence of subsets in order, and after *d* steps, we end up in the same subset.
- Example:



Structure of a periodic recurrent class. In this example, $d = 3$.

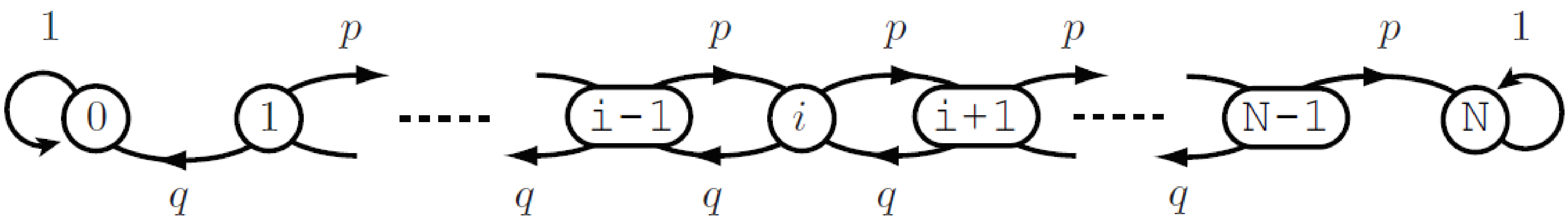
Classification of States

- **Irreducible and reducible chain**. A Markov chain with transition matrix P is **irreducible** if for any two states *i* and *j*, it is possible to go from *i* to *j* in a finite number of steps (with positive probability). That is, for any states *i*, *j* there is some positive integer *n* such that the (*i*, *j*) entry of P^n is positive. A Markov chain that is not irreducible is called **reducible**.
  - In an irreducible Markov chain with a finite state space, all states are recurrent.

Classification of States

- Example 8. Gambler's ruin as a Markov chain. Let $N \geq 2$ be an integer and let $1 \leq i \leq N - 1$. Consider a gambler who starts with an initial fortune of \$$i$ and then on each successive gamble either wins \$1 or loses \$1 independent of the past with probabilities p and $q = 1 - p$ respectively. Let X_n denote the total fortune after the nth gamble. The gambler's objective is to reach a total fortune of \$$N$, without first getting ruined (running out of money). If the gambler succeeds, then the gambler is said to *win* the game. In any case, the gambler stops playing after winning or getting ruined, whichever happens first.

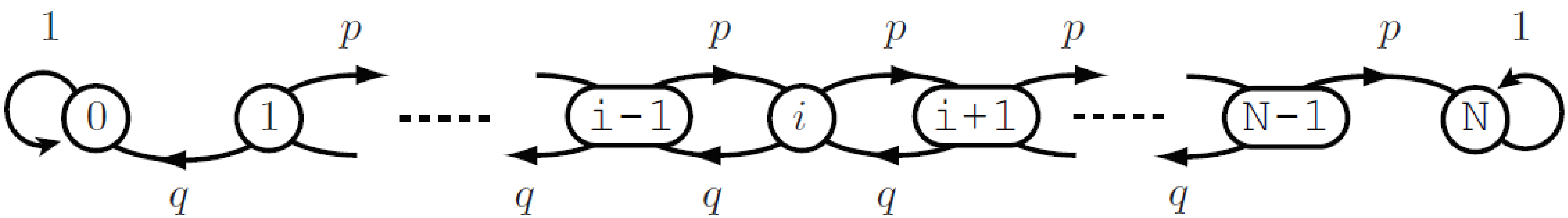
Classification of States



$\{Xn\}$ yields a Markov chain (MC) on the state space $S = \{0, 1, \ldots, N\}$. The transition probabilities are given by $P_{i,i+1} = p$; $P_{i,i\text{-}1} = q$, $0 < i < N$, and both 0 and N are absorbing states, $P_{00} = P_{NN} = 1$.

$$
P = \begin{bmatrix}
1 & 0 & 0 & . & . & . & . & . \\
q & 0 & p & 0 & . & . & . & \\
0 & q & 0 & p & 0 & . & . & \\
. & . & . & . & . & . & . & . \\
. & . & . & 0 & q & 0 & p & \\
. & . & . & . & . & . & 0 & 1
\end{bmatrix}
$$

Classification of States



Once the Markov chain reaches 0 or N, signifying bankruptcy for player A or player B, the Markov chain stays in that state forever. The probability that either A or B goes bankrupt is 1, so for any starting state *i* other than 0 or *N*, the Markov chain will eventually be absorbed into state 0 or *N*, never returning to *i*. Therefore, for this Markov chain, states 0 and *N* are recurrent, and all other states are transient. The **chain is reducible** because from state 0 it is only possible to go to state 0, and from state N it is only possible to go to state N.

Gambler's Ruin Problem Solution

- There is nothing special about starting with \$1, more generally the gambler starts with \$$i$ where $0 < i < N$.

While the game proceeds, $\{R_n : n \geq 0\}$ forms a simple random walk

$$R_n = \Delta_1 + \cdots + \Delta_n, \ R_0 = i,$$

where $\{\Delta_n\}$ forms an i.i.d. sequence of r.v.s. distributed as $P(\Delta = 1) = p, \ P(\Delta = -1) = q = 1 - p$, and represents the earnings on the succesive gambles.

Since the game stops when either $R_n = 0$ or $R_n = N$, let

$$\tau_i = \min\{n \geq 0 : R_n \in \{0, N\} | R_0 = i\},$$

denote the time at which the game stops when $R_0 = i$. If $R_{\tau_i} = N$, then the gambler wins, if $R_{\tau_i} = 0$, then the gambler is ruined.

Gambler's Ruin Problem Solution

Let $P_i = P(R_{\tau_i} = N)$ denote the probability that the gambler wins when $R_0 = i$. Clearly $P_0 = 0$ and $P_N = 1$ by definition, and we next proceed to compute P_i, $1 \le i \le N-1$.

The key idea is to condition on the outcome of the first gamble, $\Delta_1 = 1$ or $\Delta_1 = -1$, yielding

$$P_i = pP_{i+1} + qP_{i-1}. \tag{1}$$

The derivation of this recursion is as follows: If $\Delta_1 = 1$, then the gambler's total fortune increases to $R_1 = i+1$ and so by the Markov property the gambler will now win with probability P_{i+1}. Similarly, if $\Delta_1 = -1$, then the gambler's fortune decreases to $R_1 = i-1$ and so by the Markov property the gambler will now win with probability P_{i-1}. The probabilities corresponding to the two outcomes are p and q yielding (1). Since $p + q = 1$, (1) can be re-written as $pP_i + qP_i = pP_{i+1} + qP_{i-1}$, yielding

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}).$$

Gambler's Ruin Problem Solution

In particular, $P_2 - P_1 = (q/p)(P_1 - P_0) = (q/p)P_1$ (since $P_0 = 0$), so that $P_3 - P_2 = (q/p)(P_2 - P_1) = (q/p)^2 P_1$, and more generally

$$P_{i+1} - P_i = \left(\frac{q}{p}\right)^i P_1, \; 0 < i < N.$$

Thus $P_{i+1} - P_1 = (P_{i+1} - P_i) + (P_i - P_{i-1}) + (P_{i-1} - P_{i-2}) + \cdots + (P_2 - P_1)$

$$\begin{aligned} P_{i+1} - P_1 &= \sum_{k=1}^{i} (P_{k+1} - P_k) \\ &= \sum_{k=1}^{i} \left(\frac{q}{p}\right)^k P_1, \end{aligned}$$

Gambler's Ruin Problem Solution

$$
\begin{aligned}
P_{i+1} &= P_1 + P_1 \sum_{k=1}^{i} \left(\frac{q}{p}\right)^k = P_1 \sum_{k=0}^{i} \left(\frac{q}{p}\right)^k \\
&= \begin{cases} P_1 \frac{1-\left(\frac{q}{p}\right)^{i+1}}{1-\left(\frac{q}{p}\right)}, & \text{if } p \neq q; \\ P_1(i+1), & \text{if } p = q = 0.5. \end{cases}
\end{aligned} \tag{2}
$$

(Here we are using the "geometric series" equation $\sum_{n=0}^{i} a^i = \frac{1-a^{i+1}}{1-a}$, for any number a and any integer $i \geq 1$.)

Gambler's Ruin Problem Solution

Choosing $i = N - 1$ and using the fact that $P_N = 1$ yields

$$1 = P_N = \begin{cases} P_1 \frac{1-(\frac{q}{p})^N}{1-(\frac{q}{p})}, & \text{if } p \neq q; \\ P_1 N, & \text{if } p = q = 0.5, \end{cases}$$

from which we conclude that

$$P_1 = \begin{cases} \frac{1-\frac{q}{p}}{1-(\frac{q}{p})^N}, & \text{if } p \neq q; \\ \frac{1}{N}, & \text{if } p = q = 0.5, \end{cases}$$

Gambler's Ruin Problem Solution

thus obtaining from (2) (after algebra) the solution

$$P_i = \begin{cases} \frac{1-(\frac{q}{p})^i}{1-(\frac{q}{p})^N}, & \text{if } p \neq q; \\ \frac{i}{N}, & \text{if } p = q = 0.5. \end{cases}$$

(Note that $1 - P_i$ is the probability of ruin.)

Steady State Behaviour

- The concepts of recurrence and transience are important for understanding the long-run behavior of a Markov chain.
  - At first, the chain may spend time in transient states.
  - Eventually though, the chain will spend all its time in recurrent states. But what fraction of the time will it spend in each of the recurrent states?
- This question is answered by the **stationary distribution** of the chain, also known as the **steady-state** behaviour.

Steady State Behaviour

- In Markov chain models, it is interesting to determine the long-term state occupancy behaviour
  - in the *n*-step transition probabilities p_{ij} when n is very large.
- p_{ij} may converge to steady-state values that are independent of the initial state.
  - For every state j, the probability $p_{ij}(n)$ of being at state j approaches a limiting value that is independent of the initial state i, provided we exclude two situations, multiple recurrent classes/or a periodic class.

Steady State Behaviour

- This limiting value, denoted as π_j, has the interpretation:

$$\pi_j \approx \mathbf{P}(X_n = j), \qquad \text{when } n \text{ is large,}$$

- And is called the **steady-state probability** of *j*.

Steady-State Convergence Theorem

- Consider a Markov chain with a single recurrent class, which is aperiodic. Then, the states *j* are associated with steady-state probabilities π_j that have the following properties:

(a) For each *j*, we have: $\lim_{n\to\infty} p_{ij}(n) = \pi_j, \qquad \text{for all } i.$

(b) The π_j are the unique solution to the system of equations below:

$$\pi_j = \sum_{k=1}^{m} \pi_k p_{kj}, \qquad j = 1, \ldots, m,$$

$$1 = \sum_{k=1}^{m} \pi_k.$$

(c) We have:

$$\pi_j = 0, \qquad \text{for all transient states } j,$$

$$\pi_j > 0, \qquad \text{for all recurrent states } j.$$

Steady State Behaviour

- The steady-state property π_j sum to 1 and form a probability distribution on the state space, called the **stationary distribution (PMF)** of the chain.
- Thus, if the initial state is chosen according to this distribution, i.e., if:

$$\mathbf{P}(X_0 = j) = \pi_j, \qquad j = 1, \ldots, m,$$

- Then, using the total probability theorem, we have:

$$\mathbf{P}(X_1 = j) = \sum_{k=1}^{m} \mathbf{P}(X_0 = k) p_{kj} = \sum_{k=1}^{m} \pi_k p_{kj} = \pi_j,$$

Steady State Behaviour

- where the last equation follows from part (b) of the steady-state theorem.
- Similarly, we obtain $P(X_n = j) = \pi_j$, for all n and j.
- Thus, if the initial state is chosen according to the stationary distribution, the state at any future time will have the same distribution.

Steady State Behaviour

- In other words, as $n \to \infty$, the n-step transition probability matrix approaches a matrix in which all the rows are equal to the same pmf, that is,

$$p_{ij}(n) \to \pi_j \quad \text{for all } i. \tag{11.17a}$$

We can express the above in matrix notation as:

$$P^n \to \mathbf{1}\boldsymbol{\pi} \tag{11.17b}$$

where $\mathbf{1}$ is a column vector of all 1's, that is, $\mathbf{1}^{\mathrm{T}} = (1, 1, \ldots)$ and $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$. From Eq. (11.16a), the convergence of P^n implies the convergence of the state pmf's:

$$p_j(n) = \sum_i p_{ij}(n) p_i(0) \to \sum_i \pi_j p_i(0) = \pi_j. \tag{11.18}$$

We say that the system reaches "equilibrium" or "steady state."

Steady State Behaviour

We can find the pmf $\boldsymbol{\pi} \triangleq \{\pi_j\}$ in Eq. (11.18) (when it exists) by noting that as $n \to \infty$, $p_j(n) \to \pi_j$ and $p_i(n-1) \to \pi_i$, so Eq. (11.15) approaches

$$\pi_j = \sum_i p_{ij}\pi_i, \tag{11.19a}$$

which in matrix notation is

$$\boldsymbol{\pi} = \boldsymbol{\pi} P. \tag{11.19b}$$

Equation (11.19b) is underdetermined and requires the normalization equation:

$$\sum_i \pi_i = 1. \tag{11.19c}$$

We refer to $\boldsymbol{\pi}$ as the **stationary state pmf** of the Markov chain. If we start the Markov chain with initial state pmf $\mathbf{p}(0) = \boldsymbol{\pi}$, then by Eqs. (11.16b) and (11.19b) we have that the state probability vector

$$\mathbf{p}(n) = \boldsymbol{\pi} P^n = \boldsymbol{\pi} \qquad \text{for all } n.$$

Steady State Behaviour

- The equations:

$$\pi_j = \sum_{k=1}^{m} \pi_k p_{kj}, \qquad j = 1, \ldots, m,$$

are called the **balance equations**.

- Once the convergence of $p_{ij}(n)$ to some π_j is taken for granted, we can consider the equation:

$$p_{ij}(n) = \sum_{k=1}^{m} p_{ik}(n-1) p_{kj},$$

take the limit of both sides as $n \to \infty$, and recover the balance equations.

Steady State Behaviour

- Together with the normalization equation:

$$\sum_{k=1}^{m} \pi_k = 1,$$

- The balance equation can be solved to obtain the π_j .

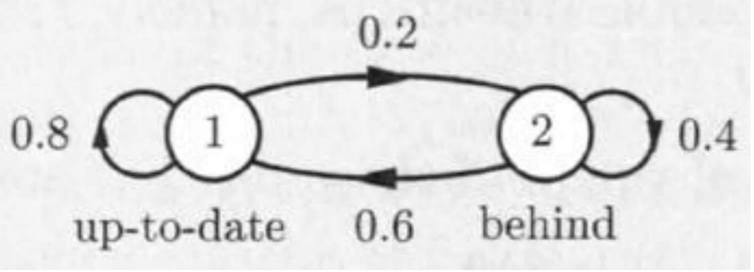
Steady State Behaviour

Example 7.1. Alice is taking a probability class and in each week, she can be either up-to-date or she may have fallen behind. If she is up-to-date in a given week, the probability that she will be up-to-date (or behind) in the next week is 0.8 (or 0.2, respectively). If she is behind in the given week, the probability that she will be up-to-date (or behind) in the next week is 0.6 (or 0.4, respectively). We assume that these probabilities do not depend on whether she was up-to-date or behind in previous weeks, so the problem has the typical Markov chain character (the future depends on the past only through the present).

$$p_{11} = 0.8, \qquad p_{12} = 0.2,$$

$$p_{21} = 0.6, \qquad p_{22} = 0.4.$$

$$\mathbf{P} = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$



Steady State Behaviour

- Find the steady-state probability of the Markov chain.
- Solution. The balance equations are:

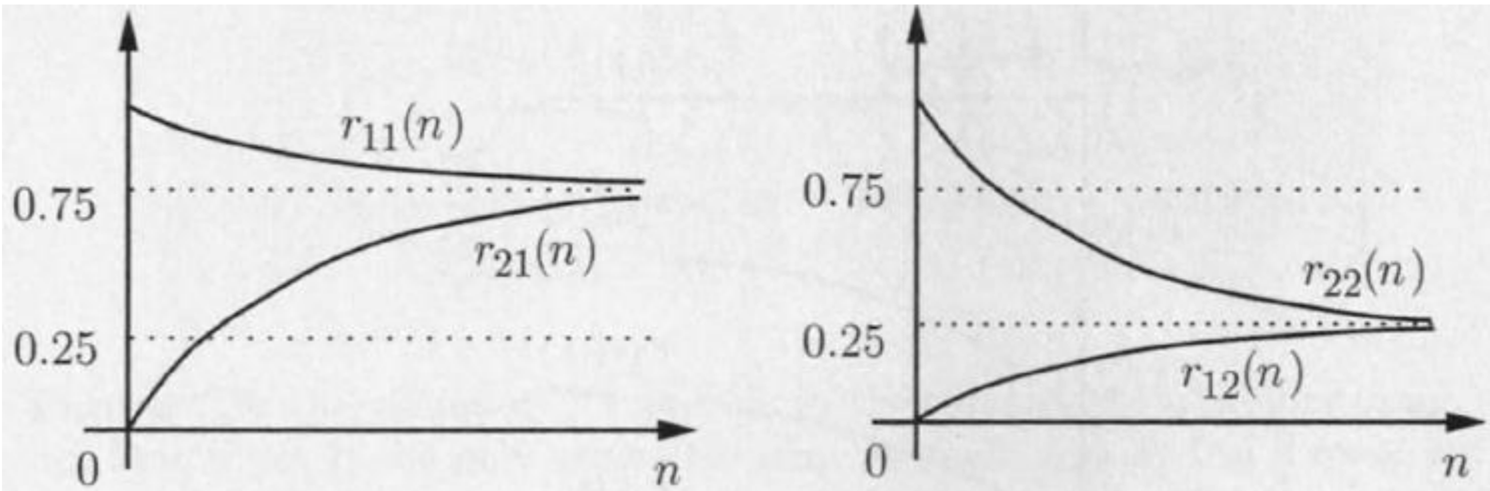
$$\pi_1 = \pi_1 p_{11} + \pi_2 p_{21}, \qquad \pi_2 = \pi_1 p_{12} + \pi_2 p_{22},$$

$$\pi_1 = 0.8 \cdot \pi_1 + 0.6 \cdot \pi_2, \qquad \pi_2 = 0.2 \cdot \pi_1 + 0.4 \cdot \pi_2.$$

$$\pi_1 = 3\pi_2. \qquad \pi_1 + \pi_2 = 1,$$

$$\pi_2 = 0.25, \qquad \pi_1 = 0.75.$$

Steady State Behaviour



n-step transition probabilities as a function of the number n of transitions

	U	B
U	0.8	0.2
B	0.6	0.4

$r_{ij}(1)$

U	.76	.24
B	.72	.28

$r_{ij}(2)$

U	.752	.248
B	.744	.256

$r_{ij}(3)$

U	.7504	.2496
B	.7488	.2512

$r_{ij}(4)$

U	.7501	.2499
B	.7498	.2502

$r_{ij}(5)$

Sequence of n-step transition probability matrices

Steady State Behaviour

- Example 2. Find the stationary distribution for the two-state Markov chain:

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

Steady State Behaviour

- Example 2. Find the stationary distribution for the two-state Markov chain:

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

- Solution: $\begin{pmatrix} s & 1-s \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} s & 1-s \end{pmatrix}$

$$\frac{1}{3}s + \frac{1}{2}(1-s) = s,$$

$$\frac{2}{3}s + \frac{1}{2}(1-s) = 1-s.$$

$s = 3/7$

$(3/7, 4/7)$ is the unique stationary distribution of the Markov chain.

Steady State Behaviour

- One way to visualize the stationary distribution of a Markov chain is to imagine a large number of particles, each independently bouncing from state to state according to the transition probabilities. After a while, the system of particles will approach an equilibrium where, at each time period, the number of particles leaving a state will be counterbalanced by the number of particles entering that state, and this will be true for all states. As a result, the system as a whole will appear to be stationary, and the proportion of particles in each state will be given by the stationary distribution.

Long-Term Frequency Interpretation

- Consider, for example, a Markov chain involving a machine, which at the end of any day can be in one of two states, working or broken down. Each time it brakes down, it is immediately repaired at a cost of $1. How are we to model the long-term expected cost of repair per day?
  - View it as the expected value of the repair cost on a randomly chosen day far into the future; this is just the steady-state probability of the broken down state.
  - Calculate the total expected repair cost in *n* days, where *n* is very large, and divide it by *n*.

Long-Term Frequency Interpretation

Steady-State Probabilities as Expected State Frequencies

For a Markov chain with a single class which is aperiodic, the steady-state probabilities π_j satisfy

$$\pi_j = \lim_{n \to \infty} \frac{v_{ij}(n)}{n},$$

where $v_{ij}(n)$ is the expected value of the number of visits to state j within the first n transitions, starting from state i.

- Based on this interpretation, π_j is the long-term expected fraction of time that the state is equal to j.
- Each time that state j is visited, there is probability p_{jk} that the next transition takes us to state k.
- We can conclude that $\pi_j p_{jk}$ can be viewed as the long-term expected fraction of transitions that move state from j to k.

Long-Term Frequency Interpretation

Expected Frequency of a Particular Transition

Consider n transitions of a Markov chain with a single class which is aperiodic, starting from a given initial state. Let $q_{jk}(n)$ be the expected number of such transitions that take the state from j to k. Then, regardless of the initial state, we have

$$\lim_{n\to\infty} \frac{q_{jk}(n)}{n} = \pi_j p_{jk}.$$

- Given the frequency interpretation of π_j and $\pi_k p_{kj}$, the balance equation:

$$\pi_j = \sum_{k=1}^{m} \pi_k p_{kj}$$

expresses the fact that the expected frequency π_j of visits to j is equal to the sum of the expected frequencies $\pi_k p_{kj}$ of transitions that lead to j.

Long-Term Frequency Interpretation

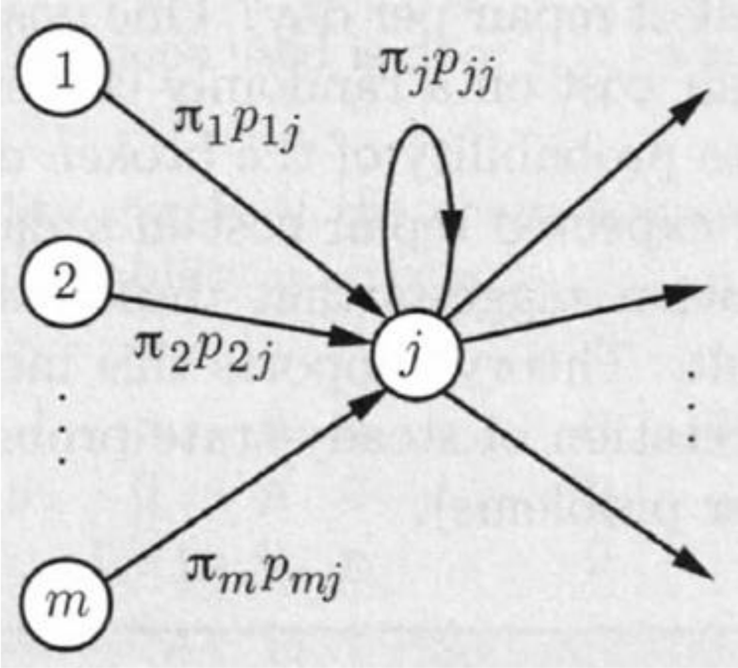
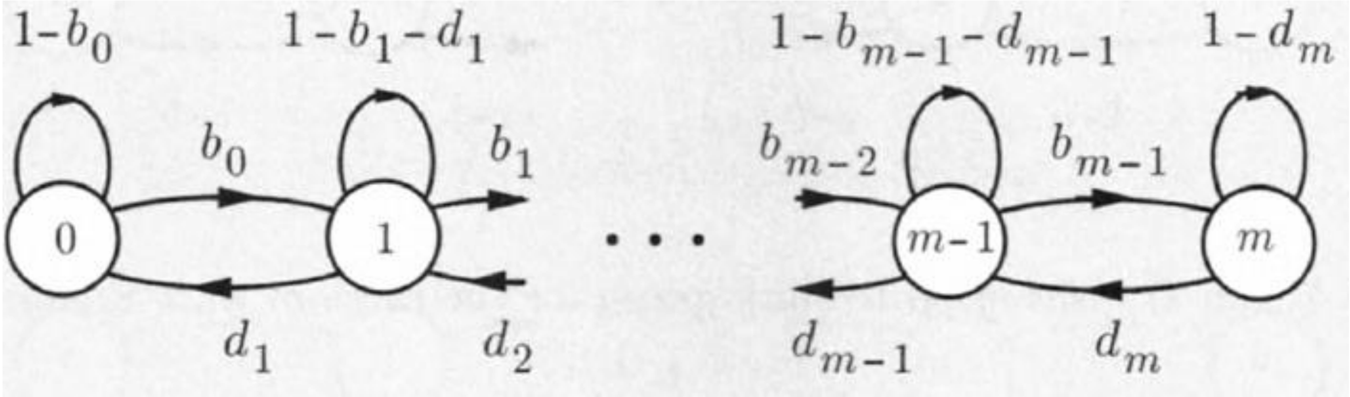


Figure 7.13: Interpretation of the balance equations in terms of frequencies. In a very large number of transitions, we expect a fraction $\pi_k p_{kj}$ that bring the state from k to j. (This also applies to transitions from j to itself, which occur with frequency $\pi_j p_{jj}$.) The sum of the expected frequencies of such transitions is the expected frequency π_j of being at state j.

Birth-Death Process

- A birth-death process is a Markov chain in which the states are linearly arranged and transitions can only occur to a neighbouring state, or else leave the state unchanged.



$$b_i = \mathbf{P}(X_{n+1} = i+1 \mid X_n = i), \qquad \text{(“birth” probability at state } i\text{)},$$
$$d_i = \mathbf{P}(X_{n+1} = i-1 \mid X_n = i), \qquad \text{(“death” probability at state } i\text{)}.$$

Birth-Death Process

- In this case the balance equation can be substantially simplified. Let focus on two neighbouring states, i and $i + 1$. In any trajectory of the Markov chain, a transition from i to $i + 1$ has to be followed by a transition from $i + 1$ to i, before another transition from i to $i + 1$ occur.
- The expected frequency transitions from i to $i + 1$, which is $\pi_i b_i$, must be equal to the expected frequency of transitions from $i + 1$ to i, which is $\pi_{i+1} d_{i+1}$. This leads to the **local balance** equations:

$$\pi_i b_i = \pi_{i+1} d_{i+1}, \qquad i = 0, 1, \ldots, m-1$$

Birth-Death Process

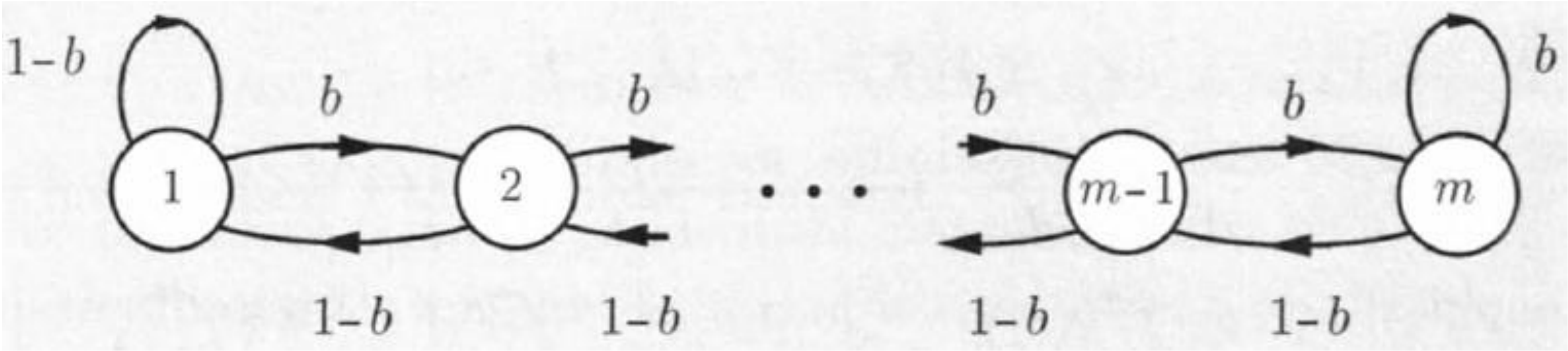
- Using the local balance equation, we obtain:

$$\pi_i = \pi_0 \frac{b_0 b_1 \cdots b_{i-1}}{d_1 d_2 \cdots d_i}, \qquad i = 1, \ldots, m.$$

- From which, using the normalization equation $\sum_i \pi_i = 1$, the steady state probabilities π_i are easily computed.

Birth-Death Process

Example 7.8. Random Walk with Reflecting Barriers. A person walks along a straight line and, at each time period, takes a step to the right with probability b, and a step to the left with probability $1 - b$. The person starts in one of the positions $1, 2, \ldots, m$, but if he reaches position 0 (or position $m + 1$), his step is instantly reflected back to position 1 (or position m, respectively). Equivalently, we may assume that when the person is in positions 1 or m, he will stay in that position with corresponding probability $1 - b$ and b, respectively. We introduce a Markov chain model whose states are the positions $1, \ldots, m$. The transition probability graph of the chain is given in Fig. 7.15.



Birth-Death Process

- The local balance equations are:

$$\pi_i b = \pi_{i+1}(1-b), \qquad i = 1, \ldots, m-1.$$

- Thus, $\pi_{i+1} = \rho\pi_i$, where:

$$\rho = \frac{b}{1-b},$$

- And we can express all the π_j in terms of π_1, as:

$$\pi_i = \rho^{i-1}\pi_1, \qquad i = 1, \ldots, m.$$

- Using the normalization equation $1 = \pi_1, + \cdots + \pi_m$, we obtain:

$$1 = \pi_1(1 + \rho + \cdots + \rho^{m-1})$$

Birth-Death Process

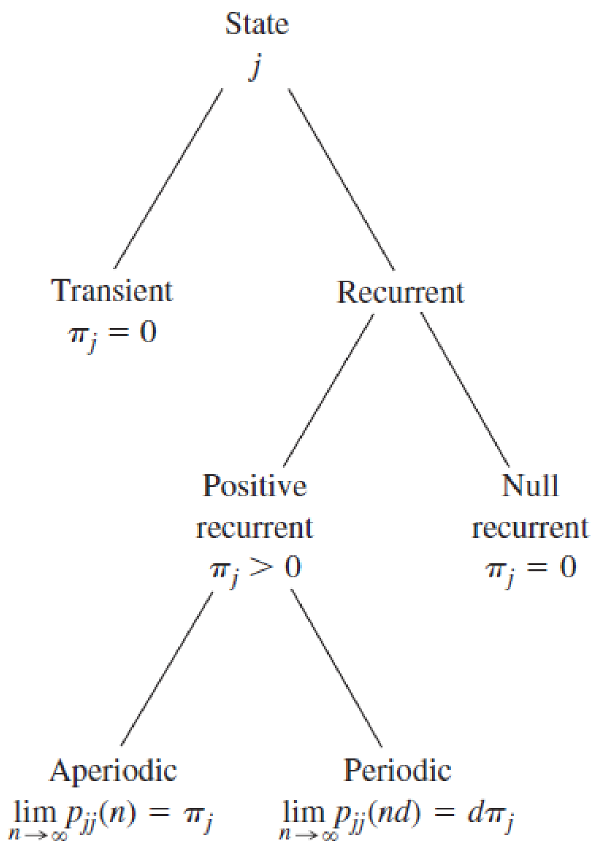
- which leads to:

$$\pi_i = \frac{\rho^{i-1}}{1 + \rho + \cdots + \rho^{m-1}}, \qquad i = 1, \ldots, m.$$

- Note that if $\rho = 1$ (left and right steps are equally likely), then $\pi_i = 1/m$ for all i.

Long-Term Frequency Interpretation

- The long-term behavior of a Markov chain is related to the types of its state classes.



Classification of states and associated long-term behavior. The proportion of time spent in state j is denoted by π_j.

Absorption Probabilities and Expected Time to Absorption

- What is the short-time behaviour of Markov chains??
  - Consider the case where the Markov chain starts at a transient state.
  - We are interested in the first recurrent state to be entered, as well as in the time until this happens.
- When addressing such questions, the subsequent behaviour of the Markov chain (after a recurrent state is encountered) is immaterial.

Absorbing Markov Chains

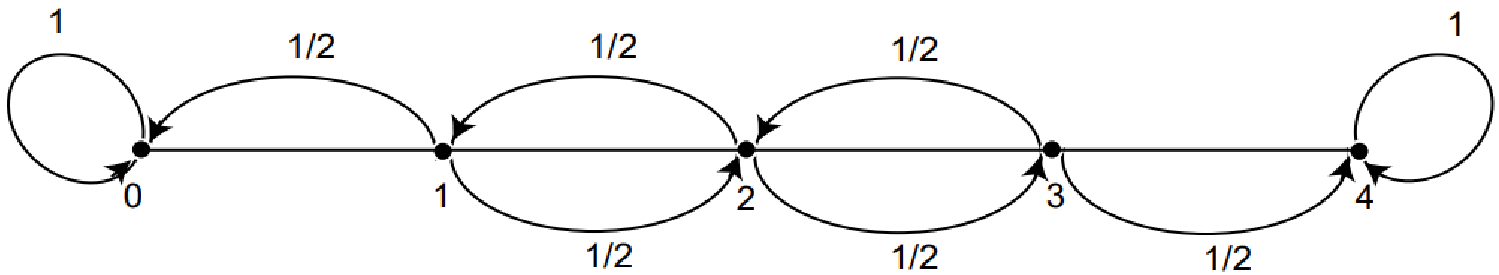


Figure 11.3: Drunkard's walk.

Definition 11.1 A state s_i of a Markov chain is called *absorbing* if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is *absorbing* if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step). □

Definition 11.2 In an absorbing Markov chain, a state which is not absorbing is called *transient*. □

Absorption Probabilities

- Focusing on the case where every recurrent state k is **absorbing**, i.e.,

$$p_{kk} = 1, \qquad p_{kj} = 0 \;\text{ for all } j \neq k$$

- If there is a unique absorbing state k, its steady-state probability is 1, and will be reached with probability 1, starting from any initial state.
  - Because all other states are transient and have zero steady-state probability.

Absorption Probabilities

- If there are multiple absorbing states, the probability that one of them will be eventually reached is still 1, but the identity of the absorbing state to be entered is random and the associated probabilities may depend on the starting state.
- Thus, we fix a particular absorbing state, denoted by *s*, and consider the absorption probability a_i that *s* is eventually reached, starting from *i*:

$$a_i = \mathbf{P}(X_n \text{ eventually becomes equal to the absorbing state } s \mid X_0 = i)$$

Absorption Probabilities

- Absorption probabilities can be obtained by solving a system of linear equations.
  - **Absorption Probability Equations**. Consider a Markov chain where each state is either transient or absorbing, and fix a particular absorbing state *s*. Then, the probabilities a_i of eventually reaching state *s*, starting from *i*, are the unique solution to the equations:

$$a_s = 1,$$

$$a_i = 0, \qquad \text{for all absorbing } i \neq s,$$

$$a_i = \sum_{j=1}^{m} p_{ij} a_j, \qquad \text{for all transient } i.$$

Absorption Probabilities

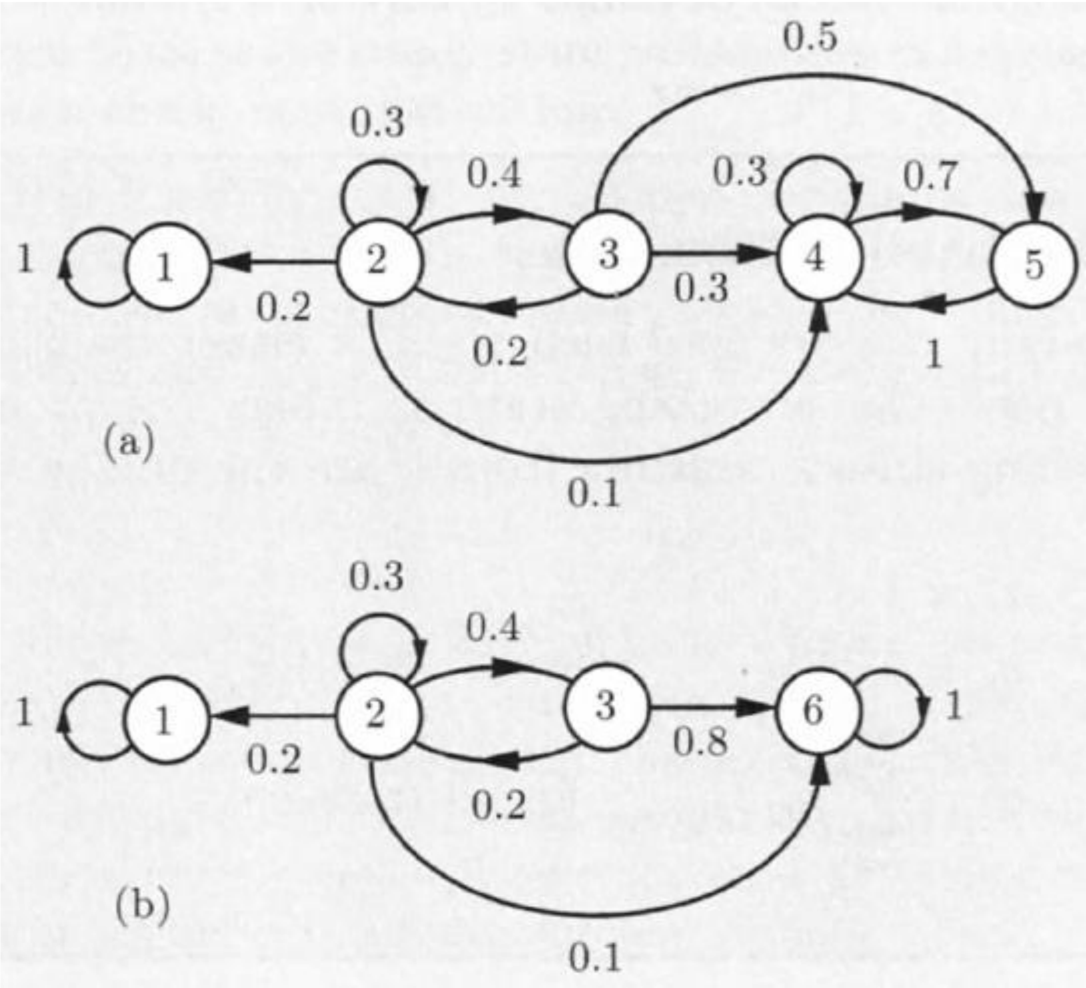
- The equations $a_s = 1$, and $a_i = 0$, for all absorbing $i \neq s$, are evident from the definition.
- The remaining equations are verified as follows:
  - Consider a transition state *i* and let *A* be the event that state *s* is eventually reached. We have:

$$
\begin{aligned}
a_i &= \mathbf{P}(A \mid X_0 = i) \\
&= \sum_{j=1}^{m} \mathbf{P}(A \mid X_0 = i, X_1 = j)\mathbf{P}(X_1 = j \mid X_0 = i) \qquad \text{(total probability thm.)} \\
&= \sum_{j=1}^{m} \mathbf{P}(A \mid X_1 = j) p_{ij} \qquad \text{(Markov property)} \\
&= \sum_{j=1}^{m} a_j p_{ij}.
\end{aligned}
$$

Absorption Probabilities

Example 7.10. Consider the Markov chain shown in Fig. 7.17(a). Note that there are two recurrent classes, namely $\{1\}$ and $\{4, 5\}$. We would like to calculate the probability that the state eventually enters the recurrent class $\{4, 5\}$ starting from one of the transient states. For the purposes of this problem, the possible transitions within the recurrent class $\{4, 5\}$ are immaterial. We can therefore lump the states in this recurrent class and treat them as a single absorbing state (call it state 6), as in Fig. 7.17(b). It then suffices to compute the probability of eventually entering state 6 in this new chain.

Absorption Probabilities



Absorption Probabilities

- The probabilities of eventually reaching state 6, starting from the transient states 2 and 3, satisfy the following equation:

$$a_2 = 0.2a_1 + 0.3a_2 + 0.4a_3 + 0.1a_6.$$

$$a_3 = 0.2a_2 + 0.8a_6.$$

- Using the fact that $a_1 = 0$ and $a_6 = 1$, we obtain:

$$a_2 = 0.3a_2 + 0.4a_3 + 0.1.$$

$$a_3 = 0.2a_2 + 0.8.$$

- Solving gives $a_2 = 21/31$ and $a_3 = 29/31$.

Expected Time to Absorption

- What is the expected number of steps until a recurrent state is entered (an event referred to as "absorption"), starting from a particular transient state?
- For any state i, we denote:

$$\mu_i = \mathbf{E}\big[\text{number of transitions until absorption, starting from } i\big]$$
$$= \mathbf{E}\big[\min\{n \geq 0 \,|\, X_n \text{ is recurrent}\} \;\big|\; X_0 = i\big].$$

- Note that if i is recurrent, then $\mu_i = 0$ according to this definition.

Expected Time to Absorption

Equations for the Expected Time to Absorption

The expected times to absorption, $\mu_1, \ldots, \mu_m$, are the unique solution to the equations

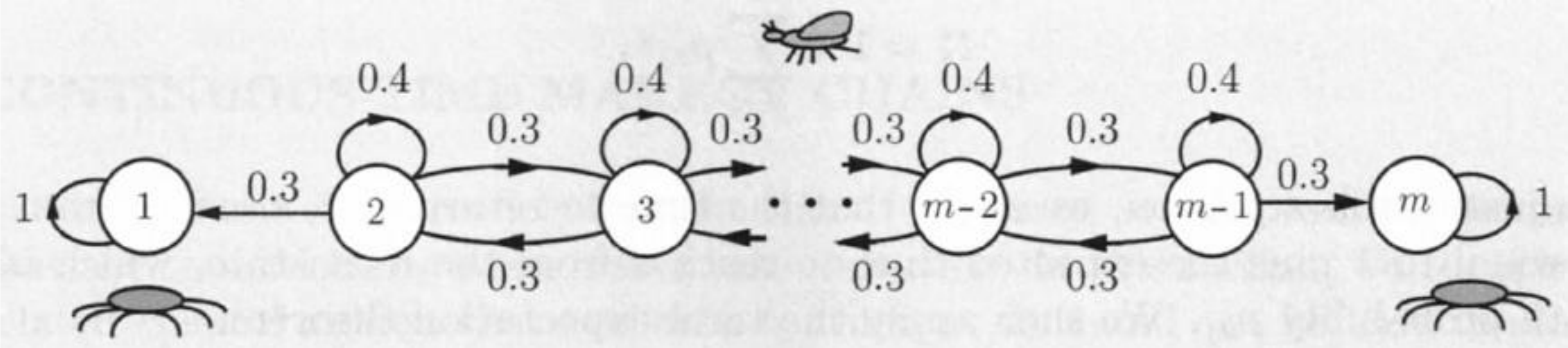
$$\mu_i = 0, \qquad \text{for all recurrent states } i,$$

$$\mu_i = 1 + \sum_{j=1}^{m} p_{ij}\mu_j, \qquad \text{for all transient states } i.$$

We argue that the time to absorption starting from a transient state i is equal to 1 plus the expected time to absorption starting from the next state, which is j with probability p_{ij}.

Expected Time to Absorption

Example 7.12. Spiders and Fly. Consider the spiders-and-fly model of Example 7.2. This corresponds to the Markov chain shown in Fig. 7.19. The states correspond to possible fly positions, and the absorbing states 1 and m correspond to capture by a spider.



- Calculate the expected number of steps until the fly is captured.

Expected Time to Absorption

- We have: $\mu_1 = \mu_m = 0,$
- And

$$\mu_i = 1 + 0.3\mu_{i-1} + 0.4\mu_i + 0.3\mu_{i+1}, \qquad \text{for } i = 2, \ldots, m-1.$$

- This equations can be solved in a variety of ways, such as for example by successive substitutions.
- As an illustration, let *m* = 4, in which case, the equations reduce to:

$$\mu_2 = 1 + 0.4\mu_2 + 0.3\mu_3, \qquad \mu_3 = 1 + 0.3\mu_2 + 0.4\mu_3.$$

- The first equation yields $\mu_2 = (1/0.6) + (1/2)\mu_3,$ which can be substituted in the second equation to give $\mu_3 = 10/3$ and by substitution again, $\mu_2 = 10/3.$

Mean First Passage and Recurrence Times

- The idea used to calculate the expected time to absorption can also be used to calculate the expected time to reach a particular recurrent state, starting from any other state.
- For simplicity, consider a Markov chain with a single recurrent class.

Mean First Passage and Recurrence Times

- Let focus on a special recurrent state *s*, and denote by t_i the **mean first passage time from state *i* to state *s***, defined by:

$$t_i = \mathbf{E}\big[\text{number of transitions to reach } s \text{ for the first time, starting from } i\big]$$
$$= \mathbf{E}\big[\min\{n \geq 0 \,|\, X_n = s\} \;\big|\; X_0 = i\big].$$

- The transitions out of state *s* are irrelevant to the calculation of the mean first passage times.

Mean First Passage and Recurrence Times

- Consider thus a new Markov chain which is identical to the original, except that the special state s is converted into an absorbing state (by setting $p_{ss} = 1$, and $p_{sj} = 0$ for all $j \neq s$).
- Whit this transformation, all states other than s become transient.

Mean First Passage and Recurrence Times

- Then, compute t_i as the expected number of steps to absorption starting from *i*, using the formulas given earlier:

$$t_i = 1 + \sum_{j=1}^{m} p_{ij} t_j, \qquad \text{for all } i \neq s,$$

$$t_s = 0.$$

- This system of linear equations can be solved for the unknowns t_i, and has a unique solution.
- These equations give the expected time to reach the special state *s* starting from any other state.

Mean First Passage and Recurrence Times

- We may also want to calculate the **mean recurrence time** of the special state *s*, which is defined as:

$$t_s^* = \mathbf{E}[\text{number of transitions up to the first return to } s, \text{ starting from } s]$$
$$= \mathbf{E}\big[\min\{n \geq 1 \,|\, X_n = s\} \;\big|\; X_0 = s\big].$$

- Then t^*_s can be obtained once we have the first passage times t_i, by using the equation:

$$t_s^* = 1 + \sum_{j=1}^{m} p_{sj} t_j .$$

Mean First Passage and Recurrence Times

- This equation can be justified saying that the time to return to s, starting from s, is equal to 1 plus the expected time to reach s from the next state, which is j with probability p_{sj}. Then apply the total expectation theorem.

Mean First Passage and Recurrence Times

Equations for Mean First Passage and Recurrence Times

Consider a Markov chain with a single recurrent class, and let s be a particular recurrent state.

- The mean first passage times t_i to reach state s starting from i, are the unique solution to the system of equations

$$t_s = 0, \qquad t_i = 1 + \sum_{j=1}^{m} p_{ij} t_j, \qquad \text{for all } i \neq s.$$

- The mean recurrence time t_s^* of state s is given by

$$t_s^* = 1 + \sum_{j=1}^{m} p_{sj} t_j.$$

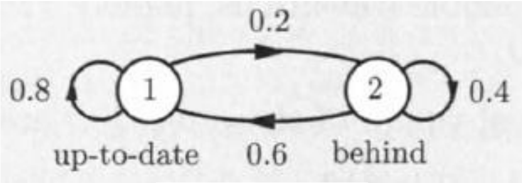
Mean First Passage and Recurrence Times

Example 7.13. Consider the "up-to-date"–"behind" model of Example 7.1. States 1 and 2 correspond to being up-to-date and being behind, respectively, and the transition probabilities are

$$p_{11} = 0.8, \qquad p_{12} = 0.2,$$
$$p_{21} = 0.6, \qquad p_{22} = 0.4.$$

Let us focus on state $s = 1$ and calculate the mean first passage time to state 1, starting from state 2. We have $t_1 = 0$ and

$$t_2 = 1 + p_{21}t_1 + p_{22}t_2 = 1 + 0.4t_2,$$



$$\begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$

from which

$$t_2 = \frac{1}{0.6} = \frac{5}{3}.$$

The mean recurrence time to state 1 is given by

$$t_1^* = 1 + p_{11}t_1 + p_{12}t_2 = 1 + 0 + 0.2 \cdot \frac{5}{3} = \frac{4}{3}.$$

Continuous Time Markov Chains

- In discrete Markov chains models it is assumed that the transitions between states take unit time.
- Continuous time Markov chains evolve in continuous time.
  - Can be used to study systems involving continuous-time arrival processes.
  - Examples: Distribution centres or nodes in communication networks where some events of interest are described in terms of Poisson processes.

Continuous Time Markov Chains

- Similar to the discrete Markov chains, continuous time Markov chains involve transitions from one state to the next:
  - According to a given transition probabilities
  - The time spend between transitions is modelled as continuous random variables.
  - It is assumed that the number of states is finite
  - In absence of a statement to the contrary, the state space is the set $S = \{1, \ldots, m\}$.

Continuous Time Markov Chains

- To describe a continuous Markov chain, some random variables of interest are introduced:

X_n : the state right after the nth transition;

Y_n : the time of the nth transition:

T_n : the time elapsed between the $(n-1)$st and the nth transition.

- For completeness, X_0 denotes the initial state, and $Y_0 = 0$.

Continuous Time Markov Chains

Continuous-Time Markov Chain Assumptions

- If the current state is i, the time until the next transition is exponentially distributed with a given parameter ν_i, independent of the past history of the process and of the next state.

- If the current state is i, the next state will be j with a given probability p_{ij}, independent of the past history of the process and of the time until the next transition.

Continuous Time Markov Chains

- The above assumptions are a complete description of the process and provide an unambiguous method for simulating it
  - Given that we just entered state *i*, we remain at state *i* for a time that is exponentially distributed with parameter v_i, and then move to a next state *j* according to the transition probabilities p_{ij}.
  - Thus, the sequence of states X_n obtained after successive transitions is a discrete-time Markov chain, with transition probabilities p_{ij}, called embedded Markov chain.

Continuous Time Markov Chains

- In mathematical terms, let:

$$A = \{T_1 = t_1, \ldots, T_n = t_n,\ X_0 = i_0, \ldots, X_{n-1} = i_{n-1},\ X_n = i\}$$

be an event that captures the history of the process until the *n*th transition.

- We then have:

$$\begin{aligned}
\mathbf{P}(X_{n+1} = j,\ T_{n+1} \geq t \mid A) &= \mathbf{P}(X_{n+1} = j,\ T_{n+1} \geq t \mid X_n = i) \\
&= \mathbf{P}(X_{n+1} = j \mid X_n = i)\,\mathbf{P}(T_{n+1} \geq t \mid X_n = i) \\
&= p_{ij} e^{-\nu_i t}, \qquad \text{for all } t \geq 0.
\end{aligned}$$

Continuous Time Markov Chains

- The expected time to the next transition is:

$$\mathbf{E}[T_{n+1} \mid X_n = i] = \int_0^\infty \tau \nu_i e^{-\nu_i \tau}\, d\tau = \frac{1}{\nu_i},$$

- So we can interpret v_i as the average number of transitions out of state i, per unit time spent at state i.
- v_i is called the **transition rate out of state** i.

Continuous Time Markov Chains

- Since only a fraction p_{ij} of the transitions out of state i will led to state j, we may also view:

$$q_{ij} = \nu_i p_{ij}$$

as the average number of transitions from i to j, per unit time spent at i.

- Thus, q_{ij} is called the **transition rate from i to j**.

Continuous Time Markov Chains

- Given the transition rates q_{ij}, one can obtain the transition rate v_i using the formula:

$$\nu_i = \sum_{j=1}^{m} q_{ij},$$

- And the transition probabilities using the formula:

$$p_{ij} = \frac{q_{ij}}{\nu_i}.$$

Continuous Time Markov Chains

- The model allows for self transitions, from a state back to itself, which can happen if a self-transition probability p_{ii} is nonzero.
- Self-transitions have no observable effects
  - Because the memorylessness of the exponential distribution, the remaining time until the next transition is the same, irrespective of whether a self-transition just occurred or not.
  - Then, self-transitions can be ignored and assume that:

$$p_{ii} = q_{ii} = 0, \qquad \text{for all } i.$$

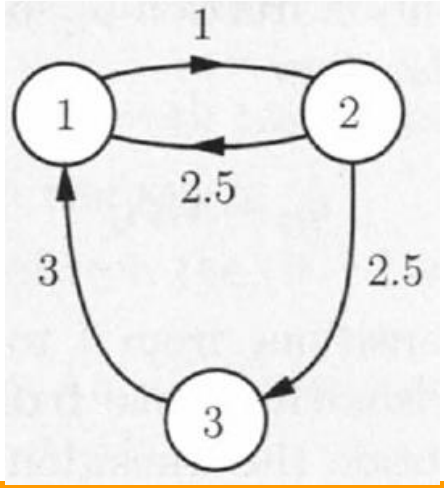
Continuous Time Markov Chains

Example 7.14. A machine, once in production mode, operates continuously until an alarm signal is generated. The time up to the alarm signal is an exponential random variable with parameter 1. Subsequent to the alarm signal, the machine is tested for an exponentially distributed amount of time with parameter 5. The test results are positive, with probability $1/2$, in which case the machine returns to production mode, or negative, with probability $1/2$, in which case the machine is taken for repair. The duration of the repair is exponentially distributed with parameter 3. We assume that the above mentioned random variables are all independent and also independent of the test results.

Continuous Time Markov Chains

Let states 1, 2, and 3, correspond to production mode, testing, and repair, respectively. The transition rates are $\nu_1 = 1$, $\nu_2 = 5$, and $\nu_3 = 3$. The transition probabilities and the transition rates are given by the following two matrices:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1 & 0 & 0 \end{bmatrix}, \qquad Q = \begin{bmatrix} 0 & 1 & 0 \\ 5/2 & 0 & 5/2 \\ 3 & 0 & 0 \end{bmatrix}.$$



Continuous Time Markov Chains

- Similar to its discrete-time counterpart, the continuous-time process has a Markov property: the future is independent of the past, given the present.

Continuous Time Markov Chains

- Approximation by a discrete-time Markov Chain
  - Let us fix a small positive number δ and consider the discrete-time Markov chain Z_n that is obtained by observing $X(t)$ every δ time units:

$$Z_n = X(n\delta), \qquad n = 0. 1, \ldots .$$

  - As Z_n is a MC, means that the future is independent from the past, given the present (The Markov property of X(t))
  - Let use $\bar{p}_{ij}$ to describe the transition probabilities of Z_n

Continuous Time Markov Chains

- Approximation by a discrete-time Markov Chain
  - Given that $Z_n = i$, there is a probability approximately equal to $\nu_i\delta$ that there is a transition between times $n\delta$ and $(n + 1)\delta$, and in that case there is a further probability p_{ij} that the next state is j:

$$\overline{p}_{ij} = \mathbf{P}(Z_{n+1} = j \mid Z_n = i) = \nu_i p_{ij}\delta + o(\delta) = q_{ij}\delta + o(\delta), \qquad \text{if } j \neq i,$$

where $o(\delta)$ is a term that is negligible compared to δ, as δ gets smaller. The probability of remaining at i [i.e., no transition occurs between times $n\delta$ and $(n+1)\delta$] is

$$\overline{p}_{ii} = \mathbf{P}(Z_{n+1} = i \mid Z_n = i) = 1 - \sum_{j \neq i} \overline{p}_{ij}.$$

Continuous Time Markov Chains

Alternative Description of a Continuous-Time Markov Chain

Given the current state i of a continuous-time Markov chain, and for any $j \neq i$, the state δ time units later is equal to j with probability

$$q_{ij}\delta + o(\delta),$$

independent of the past history of the process.

Continuous Time Markov Chains

Example 7.14 (continued). Neglecting $o(\delta)$ terms, the transition probability matrix for the corresponding discrete-time Markov chain Z_n is

$$\begin{bmatrix} 1-\delta & \delta & 0 \\ 5\delta/2 & 1-5\delta & 5\delta/2 \\ 3\delta & 0 & 1-3\delta \end{bmatrix}.$$

Continuous Time Markov Chains

- Steady-state behavior
- Birth-Death Processes

Ergodic Theorem for Discrete Markov Chains

Markov Chain Montecarlo Method

Queuing Theory