Марковские процессы

А.Г. Трофимов
к.т.н., доцент, НИЯУ МИФИ

lab@neuroinfo.ru
http://datalearning.ru

Курс "Статистическая обработка временных рядов"

Сентябрь 2018

Why Markov Processes?

Why Markov processes are important?

- Many analytical methods and solutions are developed only for Markov processes
  The usual way to solve problems in signal processing is adjust them to some Markov models
- Markov processes are adequate for many real-life phenomena
  Furthermore, some real-life processes can be approximated by Markov processes

A Markov process is a stochastic process that satisfies the Markov property, which means that the past and future are independent when the present is known

This means that if one knows the current state of the process, then no additional information of its past states is required to make the best possible prediction of its future

Formal Definition of Markov Process

Definition

The random process $\{X(t), t \in \mathscr{T}\}$ is called Markov process if for any $n > 2$ and any time moments $t_0 < t_1 < ... < t_n \in \mathscr{T}$ the conditional CDF of "last" random variable $X(t_n)$ given the fixed values of $X(t_0), ..., X(t_{n-1})$ depends only on $X(t_{n-1})$:

$$P(X(t_n) < x_n \mid X(t_{n-1}) = x_{n-1} \; \& \; ... \; \& \; X(t_0) = x_0)$$
$$= P(X(t_n) < x_n \mid X(t_{n-1}) = x_{n-1})$$

For $n = 3$ time moments $t_0 < t_1 < t_2$:

$$P(X(t_2) < x_2 \mid X(t_1) = x_1 \; \& \; X(t_0) = x_0)$$
$$= P(X(t_2) < x_2 \mid X(t_1) = x_1)$$

Examples of Markov Time Series

- IID time series

  Given $X(t)$ the random variable $X(t+\tau)$ depends neither on the past nor on $X(t)$ for all $t \in \mathscr{T}$ and $\tau > 0$

  Particularly, strongly white time series is Markov time series

- SII time series

  $\{X(t), t = 0, 1, ...\}$ is SII time series $\Leftrightarrow$ $X(t) = \sum_{i=1}^{t} U(i)$,

  where $\{U(t), t = 1, 2, ...\}$ is IID time series

  Given $X(t)$ the random variable

  $$X(t+\tau) = \sum_{i=1}^{t+\tau} U(i) = X(t) + \sum_{i=t+1}^{t+\tau} U(i)$$

  doesn't depend on the past before time t

  Particularly, Bernoulli counting process and simple random walk are Markov time series

Example of Non-Markov Time Series

Consider Gaussian time series $\{X(t), t = 1, 2, ...\}$:

$$X(t) = X(t-1) - X(t-2) + U(t)$$

$$X(1) = U(1), \quad X(2) = U(2)$$

where U is a Gaussian white noise $N(0, 1)$

For time moments $t-2, t-1, t$:

$$P(X(t) < x \mid X(t-1) = x_1 \; \& \; X(t-2) = x_2)$$
$$\neq P(X(t) < x \mid X(t-1) = x_1)$$

The distribution $X(t)|_{X(t-1)=x_1}$ depends on x_2

It is insufficient to know only the current state $X(t) = x$ to deduce statistical properties of the future random variable $X(t+1)$. It is also necessary to know the past before time moment t

Therefore the time series X is non-Markovian

Markov Property. Notes

- The Markov property is referred to as the "memorylessness" property
- For Markov process its future (i.e., the distribution of future outcomes) only depends on the current state, but not its past
- We don't need to know the full history of states to know what will happen next, just the current one
- The Markov property is desired property in predictive modelling tasks
- The Markov property leads to great reduction of the number of parameters when studying such processes
- Some non-Markovian processes can be transformed to Markov ones in high dimensional spaces

Markov Chain

Consider a process X, $X(t) \in \mathscr{X}$, $t \in \mathscr{T}$

The index set $\mathscr{T}$ and the state space $\mathscr{X}$ can be discrete (countable or finite) or continuous sets

The discrete-time process $\{X(t), t = 0, 1, ...\}$ is a sequence of random variables $X(0), X(1), ...$ so we'll denote it as $X_0, X_1, ...$, where $X_n \equiv X(n)$, $n = 0, 1, ...$

Definition

Markov chain is a discrete-time process $\{X_n, n = 0, 1, ...\}$ in discrete state space $\mathscr{X} = S \subseteq \{1, 2, ...\}$, such that

$$P(X_{n+1} = j \mid X_n = i \;\&\; X_{n-1} = i_{n-1} \;\&\; ... \;\&\; X_0 = i_0) \\ = P(X_{n+1} = j \mid X_n = i)$$

for all n and all $i_0, i_1, ..., i_{n-1}, i, j \in S$

Markov Chain. Notes

- Sometimes the state space S of Markov chain X is considered to be continuous but many applications of Markov chains employ finite or countably infinite state spaces
- Markov chain is a discrete-time Markov process (in discrete or continuous state space)
- The state space S can be set of arbitrary objects (e.g. words, moods, actions, etc.), $S = \{s_1, s_2, ...\}$. Some examples:

$$S = \{\text{sleep, eat, excercise}\}$$
$$S = \{\text{sunny, rainy}\}$$
$$S = \{\text{bear market, bull market}\}$$

- A IID time series (for example, a series of coin flips) satisfies the formal definition of a Markov chain. However, the theory is usually applied only when the probability distribution of the next step depends non-trivially on the current state

State Transition Probabilities and Diagram

Let the state space $S = \{s_1, ..., s_k\}$, where k is a number of states

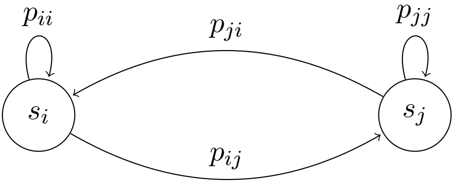
The probabilities

$$p_{ij} = P(X_{n+1} = s_j \mid X_n = s_i), \quad i, j = 1, ..., k$$

are called state transition probabilities

Transition probability p_{ij} from state s_i to state s_j depends only on i and j and doesn't depend on time step n, $i, j = 1, ..., k$

A Markov chain is usually shown by a state transition diagram



Transition Probability Matrix

Transition probabilities are arranged into the state transition matrix or transition probability matrix:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}$$

where

$$p_{ij} = P(s_j \mid s_i) = P(X_{n+1} = s_j \mid X_n = s_i), \quad i, j = 1, ..., k$$

The sum of elements in each row is equal to 1:

$$\sum_{j=1}^{k} p_{ij} = \sum_{j=1}^{k} P(s_j \mid s_i) = 1, \quad i = 1, ..., k$$

n-Step Transition Probabilities

Transition probability matrix P defines the probabilities of going from state s_i to state s_j in one step, $i, j = 1, ..., k$:

$$p_{ij} = P(X_{n+1} = s_j \mid X_n = s_i)$$

Two-step transition probabilities:

$$p_{ij}^{(2)} = P(X_{n+2} = s_j \mid X_n = s_i)$$

By the law of total probability:

$$\begin{aligned} p_{ij}^{(2)} &= \sum_{s_q \in S} P(X_{n+2} = s_j \mid X_{n+1} = s_q, X_n = s_i) P(X_{n+1} = s_q | X_n = s_i) \\ &= \sum_{s_q \in S} P(X_{n+2} = s_j \mid X_{n+1} = s_q) P(X_{n+1} = s_q | X_n = s_i) \\ &= \sum_{s_q \in S} p_{qj} p_{iq} \end{aligned}$$

Chapman-Kolmogorov Equation

Two-step transition probabilities: $p_{ij}^{(2)} = \sum_{s_q \in S} p_{iq} p_{qj}$

Two-step transition matrix: $P^{(2)} = P^2$

All n-step transition probabilities can be calculated using the Chapman-Kolmogorov equation (1931):

$$p_{ij}^{(m+n)} = P(X_{m+n} = s_j \mid X_0 = s_i) = \sum_{s_q \in S} p_{iq}^{(m)} p_{qj}^{(n)}$$

n**-step transition matrices:**

$$P^{(2)} = P^{(1)} P^{(1)} = PP = P^2$$
$$P^{(3)} = P^{(2)} P^{(1)} = P^2 P = P^3$$
$$\dots$$
$$P^{(n)} = P^n, \quad n = 1, 2, \dots$$

State Probability Distributions

Consider a Markov chain $X = \{X_n, n = 0, 1, ...\}$, where $X_n \in S = \{s_1, ..., s_k\}$

The state of the Markov chain X at step n is a discrete random variable X_n with the vector of probabilities $\pi^{(n)}$:

$$X_n \sim \pi^{(n)} = \left(\pi_1^{(n)}, ..., \pi_k^{(n)}\right) = (P(X_n = s_1), ..., P(X_n = s_k))$$

$$\pi_j^{(n)} = P(X_n = s_j) = \sum_{i=1}^{k} P(X_n = s_j \mid X_{n-1} = s_i) P(X_{n-1} = s_i)$$

$$= \sum_{i=1}^{k} p_{ij} P(X_{n-1} = s_i) = \sum_{i=1}^{k} p_{ij} \pi_i^{(n-1)}, \quad j = 1, ..., k$$

In matrix form:

$$\pi^{(n)} = \pi^{(n-1)} P$$

Stationary Distribution of Markov Chain

Assume the initial state distribution of the Markov chain is defined by probability vector $\pi^{(0)}$: $X_0 \sim \pi^{(0)}$

The probability distribution vectors of random variables $X_1, X_2, \ldots$:

$$X_1 \sim \pi^{(1)} = \pi^{(0)} P$$

$$X_2 \sim \pi^{(2)} = \pi^{(1)} P = \pi^{(0)} P^2$$

$$\ldots$$

$$X_n \sim \pi^{(n)} = \pi^{(n-1)} P = \ldots = \pi^{(0)} P^n$$

Definition

A stationary distribution π of a Markov chain is a state probability distribution that remains unchanged in the Markov chain as time progresses:

$$\pi = \pi P$$

Stationary Markov Chain

Definition

A stationary Markov chain is Markov chain with stationary state distribution π:

$$\pi = \pi^{(0)} = \pi^{(1)} = ... = \pi^{(n)} = ...$$

The stationary state distribution π of the Markov chain with state space $S = \{s_1, ..., s_k\}$ and transition probability matrix P is a solution of the problem:

$$\begin{cases} \pi = \pi P \\ \sum\limits_{j=1}^{k} \pi_j = 1 \end{cases} \Rightarrow \begin{cases} P^T \pi^T = \pi^T \\ \sum\limits_{j=1}^{k} \pi_j = 1 \end{cases}$$

The solution π^T is eigenvector of the transposed transition matrix P^T that corresponds to eigenvalue 1

Limiting Distribution of Markov Chain

Definition

The probability distribution $\pi = (\pi_1, ..., \pi_k)$ is called a limiting distribution of the Markov chain $X = \{X_n, n = 0, 1, ...\}$ with state space $S = \{s_1, ..., s_k\}$ if

$$\pi_j = \lim_{n\to\infty} P(X_n = s_j \mid X_0 = s), \quad j = 1, ..., k$$

for all $s \in S$

By definition, when a limiting distribution exists, it does not depend on the initial state $X_0 = s \in S$, so the limiting distribution of the Markov chain is a steady state probability distribution at infinity:

$$\pi_j = \lim_{n\to\infty} P(X_n = s_j) = \lim_{n\to\infty} \pi_j^{(n)}$$

Example 1. Markov Chain Description

Consider a switch that has two states: on and off. At the beginning of the experiment, the switch is off. Every minute after that, we throw a dice. If the dice shows "6", we flip the switch, otherwise we leave it as it is

The discrete-time process $\{X_n, n = 0, 1, ...\}$ of switch's states with the state space $S = \{0,1\}$ (0: off, 1: on) is a Markov chain

If we know the state of the switch at time n, we can predict its future (i.e., the distribution of states) for all future times, without requiring any knowledge about the past states

Example of non-Markov process:
We flip the switch only if the dice shows a "6" but didn't show a "6" the previous throw

Example 1. Transition Probabilities

Formula of total probability:

$$\pi_j^{(n)} = P(X_n = j) = P(X_n = j \mid X_{n-1} = 1)P(X_{n-1} = 1) + P(X_n = j \mid X_{n-1} = 0)P(X_{n-1} = 0), \quad j \in \{0, 1\}$$

$$\pi_0^{(n)} = P(X_n = 0) = \frac{1}{6}P(X_{n-1} = 1) + \frac{5}{6}P(X_{n-1} = 0)$$

$$\pi_1^{(n)} = P(X_n = 1) = \frac{5}{6}P(X_{n-1} = 1) + \frac{1}{6}P(X_{n-1} = 0)$$

Initial probabilities:

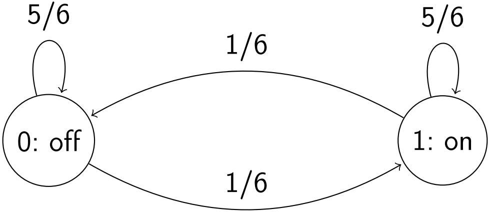
$$\pi_0^{(0)} = P(X_0 = 0) = 1$$

$$\pi_1^{(0)} = P(X_0 = 1) = 0$$

$$\pi^{(0)} = \left(\pi_0^{(0)}, \pi_1^{(0)}\right) = (1, 0)$$

Example 1. State Transition Matrix and Diagram

State transition diagram:



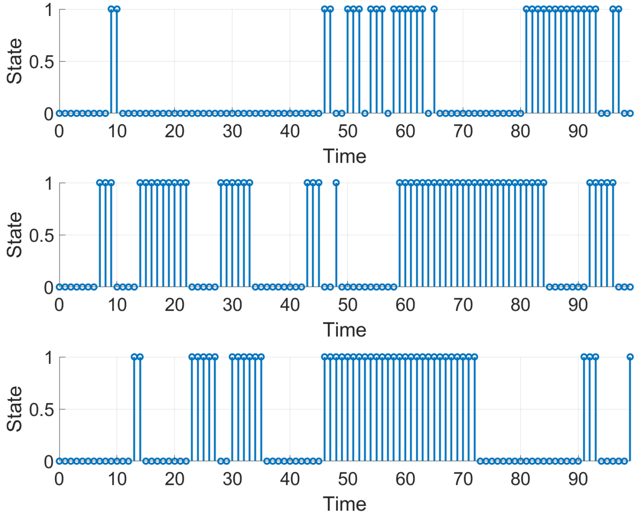
Transition probability matrix:

$$P = \begin{pmatrix} 5/6 & 1/6 \\ 1/6 & 5/6 \end{pmatrix} \approx \begin{pmatrix} 0.83 & 0.17 \\ 0.17 & 0.83 \end{pmatrix}$$

n-step transition matrices:

$$P^{(2)} \approx \begin{pmatrix} 0.72 & 0.28 \\ 0.28 & 0.72 \end{pmatrix}, \; P^{(3)} \approx \begin{pmatrix} 0.65 & 0.35 \\ 0.35 & 0.65 \end{pmatrix}, \; P^{(\infty)} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

Example 1. Sample Paths



Example 1. Limiting and Stationary Distributions

The n-step transition matrix at infinity: $P^{(\infty)} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$

It means that all state probabilities at $n \to \infty$ are equal:

$$\begin{aligned} P(X_n = 0 \mid X_0 = 0) &= P(X_n = 1 \mid X_0 = 0) \\ &= P(X_n = 0 \mid X_0 = 1) = P(X_n = 1 \mid X_0 = 1) = 0.5 \end{aligned}$$

regardless the initial distribution $\pi^{(0)}$

Therefore, the distribution $\pi = (0.5, 0.5)$ is a limiting distribution

$$\pi P = (0.5, 0.5) \begin{pmatrix} 5/6 & 1/6 \\ 1/6 & 5/6 \end{pmatrix} = (0.5, 0.5) = \pi$$

Therefore, the distribution $\pi = (0.5, 0.5)$ is a stationary distribution

Relation Between Limiting and Stationary Distributions

- Any limiting distribution π is a stationary distribution

  For limiting distribution π:

$$\pi = \lim_{n\to\infty} \pi^{(n)} = \lim_{n\to\infty} \pi^{(0)} P^n$$

$$\begin{aligned} \pi &= \lim_{n\to\infty} \pi^{(n+1)} = \lim_{n\to\infty} \pi^{(0)} P^{n+1} \\ &= \lim_{n\to\infty} \pi^{(0)} P^n P = \left(\lim_{n\to\infty} \pi^{(0)} P^n\right) P = \pi P \end{aligned}$$

  Therefore π is a stationary distribution

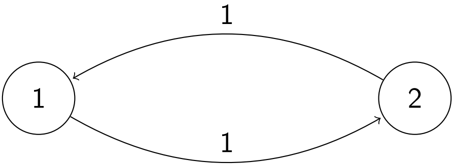
- Not all stationary distributions π are limiting distributions

  Not all Markov chains have a well-defined limiting behaviour that does not depend on the initial state probability distribution $\pi^{(0)}$

π is limiting distribution $\Rightarrow$ π is stationary distribution. The converse is not necessarily true

Example 2. Periodic Markov Chain

Consider the two-state Markov chain $X = \{X_n, n = 0, 1, ...\}$:



The transition matrix: $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

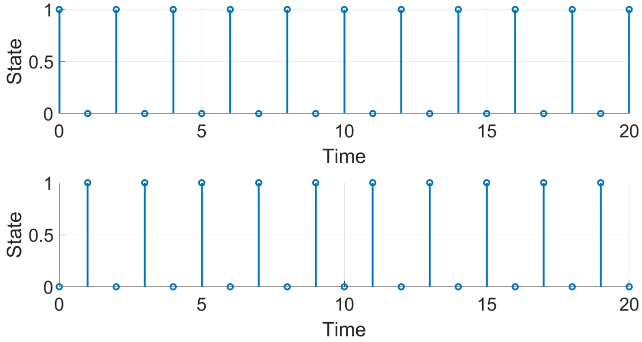
The state distribution after n steps:

$$\pi^{(n+1)} = \pi^{(n)} P = \left(\pi_1^{(n)}, \pi_2^{(n)}\right) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \left(\pi_2^{(n)}, \pi_1^{(n)}\right)$$

There is no steady state limiting distribution which doesn't depend on initial distribution $\pi^{(0)}$ for this Markov chain

Example 2. Sample Paths

The Markov chain X has only two sample paths



The state X_n at step n depends deterministically on initial state X_0 and the distribution $\pi^{(n)}$ of random variable X_n depends deterministically on initial distribution $\pi^{(0)}$

Example 2. Stationary Distribution

The stationary distribution is a solution of the system of linear equations:

$$\begin{cases} \pi = \pi P \\ \sum_{j=1}^{k} \pi_j = 1 \end{cases} \quad \Rightarrow \quad \begin{cases} (\pi_1, \pi_2) = (\pi_1, \pi_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (\pi_2, \pi_1) \\ \pi_1 + \pi_2 = 1 \end{cases}$$

The solution: $\pi_1 = \pi_2 = 0.5$

Therefore, the distribution $\pi = (0.5, 0.5)$ is a stationary distribution

If initial distribution $\pi^{(0)} = \pi = (0.5, 0.5)$, then it never will change at next time steps:

$$\pi^{(0)} = \pi^{(1)} = ... = \pi^{(n)} = (0.5, 0.5)$$

therefore, the Markov chain X is stationary

Example 3. Stationary Distribution

A sports broadcaster wishes to predict how many Michigan residents prefer University of Michigan teams and how many prefer Michigan State teams. She noticed that, year after year, most people stick with their preferred team; however, about 3% of Michigan University fans switch to Michigan State, and and about 5% of Michigan State fans switch to Michigan University. However, there is no noticeable difference in the state's population of 10 million's preference at large. What might that be? Michigan sports fans have reached a stationary distribution

Suppose x is the number of Michigan University fans and y is the number of Michigan State fans and the state population is 10 million

After one year:

$$\begin{cases} x = 0.97x + 0.05y \\ y = 0.03x + 0.95y \\ x + y = 10 \end{cases} \quad \Rightarrow \quad \begin{cases} x = 6.25 \\ y = 3.75 \end{cases}$$

$\pi = (0.625, 0.375)$ is a stationary distribution of this Markov chain

Accessible and Communicating States

Definition

The state s_j is accessible (reachable) from state s_i, written as $s_i \to s_j$, if

$$P(X_n = s_j \mid X_0 = s_i) > 0$$

for some time step $n = 0, 1, \ldots$

Definition

Two states s_i ans s_j are communicate, written as $s_i \leftrightarrow s_j$, if they are accessible from each other:

$$s_i \leftrightarrow s_j \Leftrightarrow s_i \to s_j \text{ and } s_j \to s_i$$

The communication is reflexive, symmetric and transitive relation (i.e. equivalence relation)

Communicating Classes and Irreducibility

The state set $S = \{s_1, ..., s_k\}$ can be divided into communicating classes (equivalence classes) such that only members of the same class communicate with each other

Two states s_i and s_j belong to the same communicating class if and only if $s_i \leftrightarrow s_j$

Definition

A Markov chain is said to be irreducible if all its states communicate $S = \{s_1, ..., s_k\}$ with each other, i.e. there exists a chain of steps between any two states that has positive probability

Irreducible Markov chain is Markov chain that has only one communicating class

For reducible Markov chain the asymptotic analysis is reduced to individual subclasses

Transient and Recurrent States

Transience and recurrence describe the likelihood of a process beginning in some state of returning to that particular state

Definition

The state s_i of Markov chain X is called recurrent if for some time step $n \geq 1$

$$P(X_n = s_i \mid X_0 = s_i) = 1$$

Otherwise, the state s_i is called transient

Any time that we leave the recurrent state we will return to this state in the future with probability one. For transient states the probability of returning is less than one

In communicating classes all states are recurrent or transient

A communicating class is called recurrent if the states in it are recurrent. If the states are transient, the class is called transient

Criteria of Transience and Recurrence

For transient state s_i there is a positive probability that the chain will never return to s_i after leaving it

Criterion of transience:

$$s_i \text{ is transient} \Leftrightarrow \lim_{n\to\infty} p_{ji}^{(n)} = 0 \quad \forall j = 1, ..., k$$

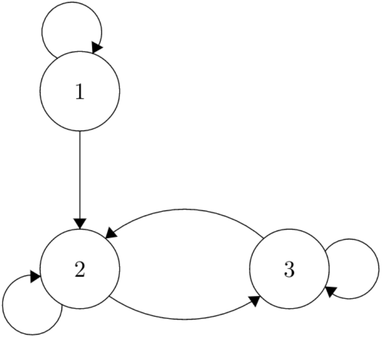
where $p_{ji}^{(n)} = (P^n)_{ji}$ is n-step transition probability from s_j to s_i

Recurrent state s_i will be visited in future with probability 1

Criterion of recurrence:

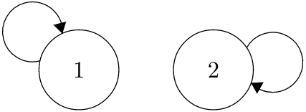
$$s_i \text{ is recurrent} \Leftrightarrow \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

Transient and Recurrent States. Examples



The state 1 is transient; the probability to return to it is less than 1

The states 2 and 3 are recurrent; if you start in either one, you you'll return with probability 1



The states 1 and 2 are recurrent; if you start in either one, you you'll return with probability 1

Positive Recurrent and Null Recurrent States

The Markov chain is expected to return to the recurrent state at $n \to \infty$ with probability 1, but it is not necessarily expected to return even once within a finite number of steps n

Definition

Let s_i be a recurrent state of Markov chain X, $X_0 = s_i$ and N_i is the number of transitions needed to return to state s_i. The state s_i is called positive recurrent if $\mathrm{M}[N_i] < \infty$. Otherwise, the state s_i is called null recurrent

For a positive recurrent state the expected number of steps to return to it is finite, and for null-recurrent is infinite

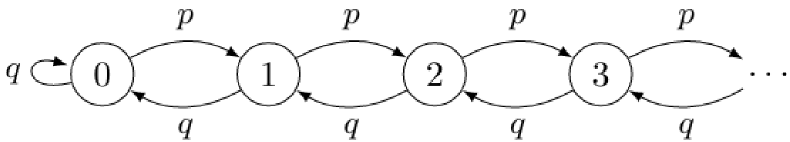
If all states in an irreducible Markov chain are positive recurrent, then the Markov chain is called positive recurrent. If all states are null recurrent, then the Markov chain is called null recurrent

Positive Recurrent and Null Recurrent States. Example

Consider a random walk with reflection at zero. It is a infinite state discrete-time Markov chain with state space $S = \{0, 1, ...\}$

Transition probabilities:

$$p_{ij} = \begin{cases} p, & j = i + 1 \\ q = 1 - p, & j = i - 1 \\ 0, & \text{otherwise} \end{cases}$$



$p < 0.5 \Rightarrow$ all states are positive recurrent
$p = 0.5 \Rightarrow$ all states are null recurrent
$p > 0.5 \Rightarrow$ all states are transient

Periodic and Aperiodic States

Definition

A state s_i has period $r \geq 1$ if any chain starting at and returning to state s_i with positive probability must take a number of steps divisible by r. If $r = 1$, then the state s_i is called aperiodic, and if $r > 1$, the state s_i is called periodic

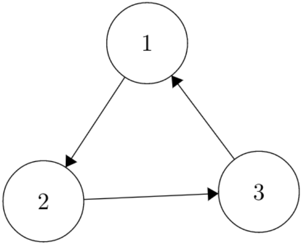
The period r of state s_i is the greatest common denominator of the lengths of all return trips, given that you start in the state s_i

All states in the same communicating class have the same period

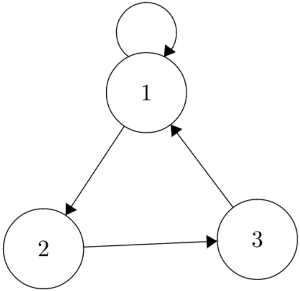
A communicating class is said to be periodic if its states are periodic. A communicating class is said to be aperiodic if its states are aperiodic

If all states in Markov chain are aperiodic, then the Markov chain is called aperiodic

Periodic and Aperiodic States. Examples



All states are periodic with period $r = 3$



All states are aperiodic (period $r = 1$);
lengths of return trips:
to state 1 are 1,2,3,...
to states 2 and 3 are 3,4,5,...

Absorbing States

A common type of Markov chain with transient states is an absorbing one

Definition

An absorbing (terminal) state is a state s_i in a Markov chain X such that

$$P(X_{n+1} = s_i \mid X_n = s_i) = 1$$

Once the absorbing state reached it is impossible to leave it

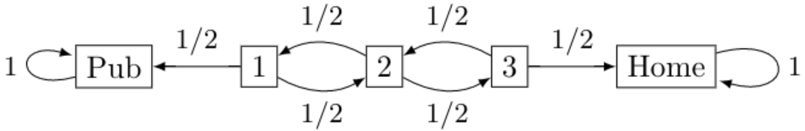
An absorbing Markov chain is a Markov chain in which it is impossible to leave some states, and any state could (after some number of steps, with positive probability) reach such a state

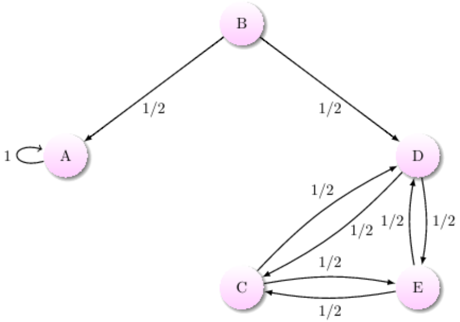
It is not sufficient for a Markov chain to contain an absorbing states to be an absorbing Markov chain

All non-absorbing states in an absorbing Markov chain are transient

Absorbing States. Examples

The Pub and Home are absorbing states of the drunkard's walk:





The state A is absorbing state but the Markov chain is not absorbing one; it is impossible to reach the absorbing state with probability 1 from all other states

Ergodic Markov Chain

Definition

An ergodic Markov chain is irreducible and aperiodic Markov chain, all states of which are positive recurrent

Ergodic Markov chains are, in some senses, the processes with the "nicest" behaviour

If a finite state irreducible Markov chain has a recurrent aperiodic state then it is ergodic

A Markov chain is ergodic if there is a number N such that any state can be reached from any other state in any number of steps greater than or equal to a number N (in case of a fully connected transition matrix $N = 1$)

By changing one state in an ergodic Markov chain into an absorbing state, the chain immediately becomes an absorbing one

Stationary Distribution of Ergodic Markov Chain

Theorem

An irreducible Markov chain has a stationary distribution if and only if all its states are positive recurrent, and in this case the stationary distribution is unique. The stationary distribution of ergodic Markov chain is unique

Theorem

Let X be an ergodic Markov chain with state space $S = \{s_1, ..., s_k\}$, initial state $X_0 = s_j$ and N_j is the number of steps needed to return to state s_j. Then the expected number of steps:

$$\mathrm{M}[N_j] = \frac{1}{\pi_j}, \quad j = 1, ..., k$$

where $\pi = (\pi_1, ..., \pi_k)$ is a stationary distribution of X

Total Variation Distance

Ergodic Markov chain has a unique stationary distribution π

How much time it is necessary the state probability distribution $\pi^{(n)}$ after n steps to be "close" to the stationary distribution π?

We need some measure between these distributions

Definition

Total variation distance $\Delta(p, q)$ between discrete distributions p and q on the same sample space Ω is defined as

$$\Delta(p, q) = \frac{1}{2} \sum_{\omega \in \Omega} |p(\omega) - q(\omega)|$$

Mixing Time of Ergodic Markov Chain

Definition

The mixing time of the Markov chain is the average time (number of steps) for the total variation distance $\Delta\left(\pi^{(0)}, \pi\right)$ between an arbitrary initial state distribution $\pi^{(0)}$ and stationary distribution π to decay by a factor of e

The mixing means the forgetting of initial state

Mixing time is a measure of the relative connectivity of transition structures in different chains

Less mixing time means more connected structure in the Markov chain and more quick approaching to the stationary distribution

The mixing time is related to the eigenvalues of transition matrix P

Perron-Frobenius Theorem

Markov chain eigenvalues are eigenvalues of its transition probability matrix

Perron-Frobenius Theorem

A Markov chain with a single recurrent aperiodic communicating class has exactly one eigenvalue equal to 1 (the Perron-Frobenius eigenvalue). All other eigenvalues have modulus less than or equal to 1. The inequality is strict unless the recurrent class is periodic

The mixing time for ergodic Markov chain:

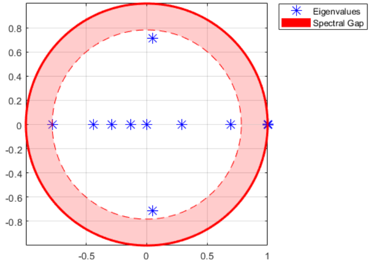
$$t_{mix} = -\frac{1}{\ln \mu}$$

where μ is second largest chain's eigenvalue magnitude

Spectral gap is the difference between first and second largest eigenvalue magnitudes: $1 - \mu$

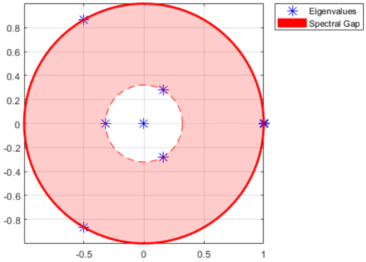
Spectral Gaps. Illustrations

For aperiodic chain



Thin spectral gap indicate slower mixing

For periodic chain

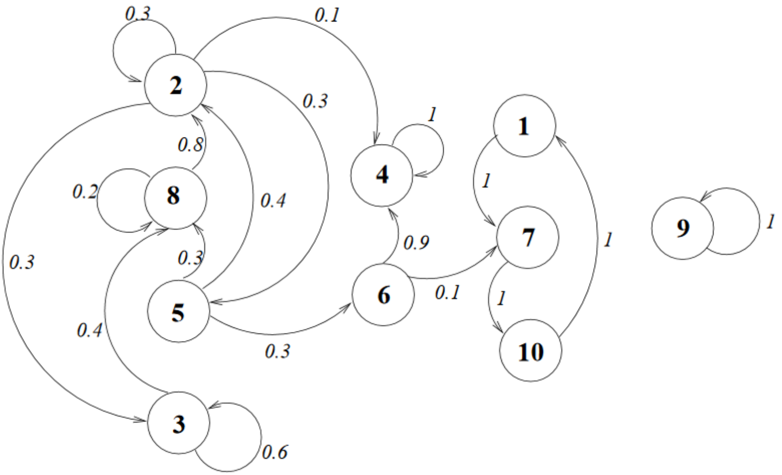


Three eigenvalues have modulus one, which indicates that the period of Markov chain is three. Large spectral gap indicate quick mixing

Ergodic Markov Chains. Notes

- Markov chain is ergodic $\Leftrightarrow$ it has stationary distribution
- For ergodic Markov chain the initial state distribution $\pi^{(0)}$ will be forgotten and its state distribution converges to unique stationary distribution $\pi = (\pi_1, ..., \pi_k)$
- For ergodic Markov chain the proportion of time spent in state s_j will converge to π_j, $j = 1, ..., k$, as time step $n \to \infty$
- We can use time averaging to estimate the state probabilities $\pi_1, ..., \pi_k$
- Ergodic Markov chains are irreducible (and therefore all states communicate to each other state)
- Sometimes irreducible Markov chain is referred to as ergodic Markov chain: the Markov chain is ergodic if it is possible to eventually get from every state to every other state with positive probability (both periodic or aperiodic)

Example. State Transition Diagram



Example. Properties of the Markov Chain

- Reducibility: the chain is reducible
  It contains more than one communicating class
- Communicating classes:
  $\{1, 7, 10\}$ – recurrent, periodic ($r = 3$)
  $\{4\}$ – recurrent, aperiodic
  $\{9\}$ – recurrent, aperiodic
  $\{2, 3, 5, 8\}$ – transient, aperiodic
  $\{6\}$ – transient, aperiodic
- Absorbing states: $\{4, 9\}$
  But the chain is not absorbing because of absorbing states cannot be reached from other states with probability 1
- Ergodicity: the chain is not reducible therefore it is not ergodic and doesn't have unique stationary distribution
- Limiting distribution: the asymptotic behaviour depends on initial state, there is no limiting distribution

Markov Chain Modelling Workflow

- **Step 1. Markov chain creation**
  - Using transition matrix P
  - Using matrix of observed transition counts C
    The transition matrix P is the normalized matrix C (such as sum in each row is equal to 1)
  - Specifying the mixing structure
    The transition matrix P is random, but number of non-zero elements, some transition probabilities, etc. are specified (appropriate if you have less specific information on a process under modelling)
- **Step 2. Markov chain visualization**
  - Draw transition state diagram
  - Use methods of graph theory to visualize

Markov Chain Modelling Workflow

- **Step 3. Identification of class structure**
  - Determine the communicating classes of the chain
  - Determine the communicating class to which each state $s_1, ..., s_k$ belongs
  - Determine whether each communicating class is recurrent or transient
  - Determine ergodicity and period of each class. The class is periodic if it is irreducible and non-ergodic
  - The recurrent class can be condensed to recurrent subchain and viewed as a "supernode"
- **Step 4. Asymptotic analysis of ergodic classes**
  - Determine stationary distribution π
  - Plot Markov chain eigenvalues (eigenvalues of transition matrix P) in complex plane
  - Determine mixing time

Markov Chain Modelling Workflow

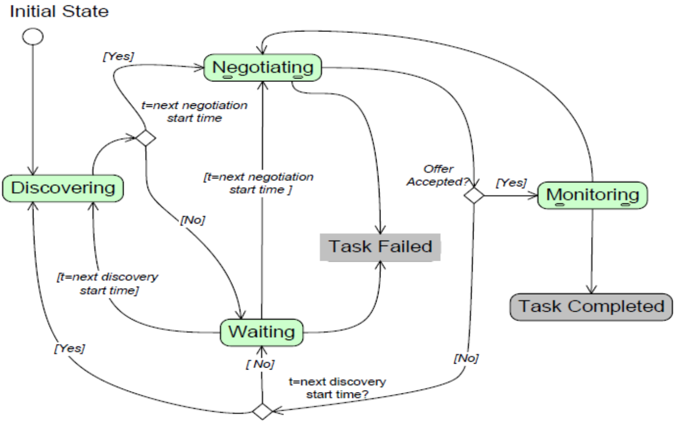
- **Step 5. Simulation and redistribution**
  Simulation provides multiple realizations of Markov chain trajectories from a specified initial state x_0 or distribution $\pi^{(0)}$

  Simulation is used to generate statistical information on the chain (by ensemble averaging of statistics) that is difficult to derive directly from the theory

  Redistribution is a calculating of n-step state distributions $\pi^{(n)}$ of Markov chain as it evolves from a specified initial distribution $\pi^{(0)}$
- **Step 6. Conclusions and model corrections**
  Make conclusions based on class structure, stationary distribution, mixing time, etc. about the process under modelling or correct the parameters of the Markov model

Markov Chain Models. Example

State model of grid compute economy simulation model*



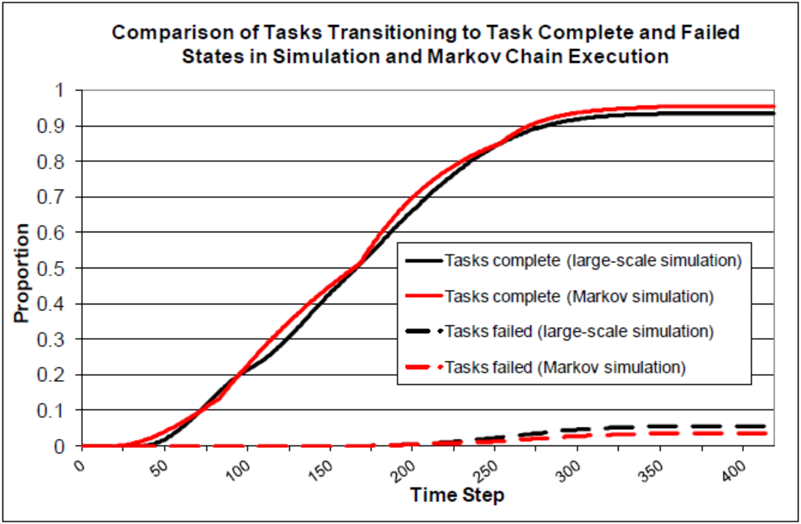
*C.Dabrowski, F.Hunt. (2009). Markov chain analysis for large-scale grid systems. National Institute of Standards and Technology (NISTIR 7566), 1-59.

Example. Transition Probability Matrix

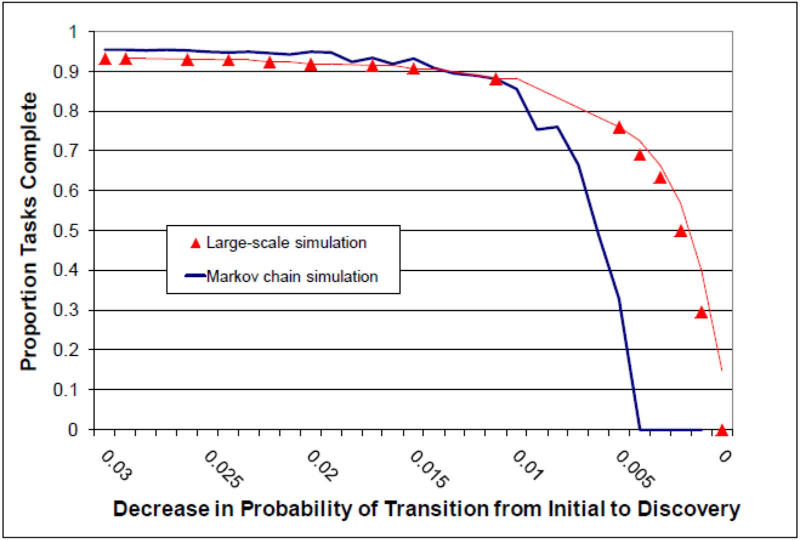
Transition probability matrix

	Initial	Wait	Disc	Ngt	Mon	Comp	Fail
Initial	0.9697	0	0.0303	0	0	0	0
Waiting	0	0.8363	0.0673	0.0918	0	0	0.0046
Disc	0	0.0355	0.6714	0.2931	0	0	0
Ngt	0	0.4974	0.0182	0.2882	0.1961	0	0.0001
Mon	0	0	0	0.0003	0.9917	0.0080	0
Comp	0	0	0	0	0	1.0	0
Fail	0	0	0	0	0	0	1.0

Example. Simulation Results 1



Example. Simulation Results 2



Example. Simulation Results 3