

# Markov chain

From Wikipedia, the free encyclopedia

In probability theory and statistics, a **Markov chain** or **Markoff chain**, named after the Russian mathematician Andrey Markov, is a stochastic process that satisfies the Markov property (usually characterized as "memorylessness"). Loosely speaking, a process satisfies the Markov property if one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history. i.e., conditional on the present state of the system, its future and past are independent.<sup>[1]</sup>

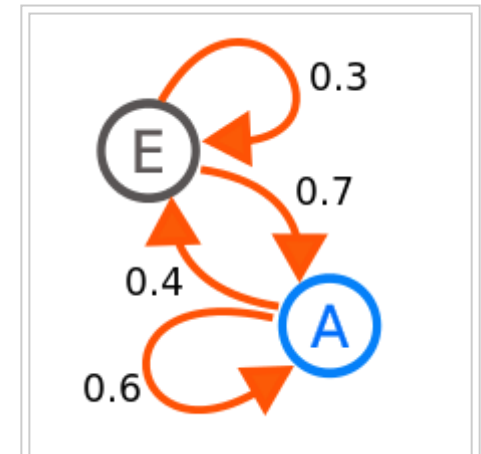
In discrete time, the process is known as a **discrete-time Markov chain** (DTMC<sup>[2]</sup>). It undergoes transitions from one state to another on a state space, with the probability distribution of the next state depending only on the current state and not on the sequence of events that preceded it.

In continuous time, the process is known as a **Continuous-time Markov chain** (CTMC<sup>[1]</sup> or **continuous-time Markov process**<sup>[2]</sup>). It takes values in some finite state space and the time spent in each state takes non-negative real values and has an exponential distribution. Future behaviour of the model (both remaining time in current state and next state) depends only on the current state of the model and not on historical behaviour.

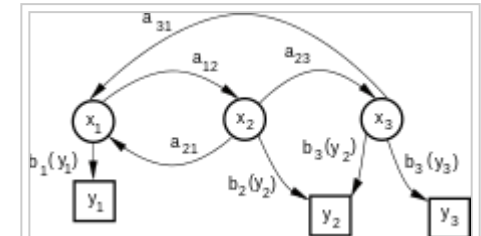
Markov chains have many applications as statistical models of real-world processes.

## Contents

- 1 Introduction
- 2 Examples
  - 2.1 Gambling
  - 2.2 A birth-death process
  - 2.3 A non-Markov example
- 3 Markov property
  - 3.1 The general case
  - 3.2 For discrete-time Markov chains
- 4 Formal definition
  - 4.1 Discrete-time Markov chain
    - 4.1.1 Variations
    - 4.1.2 Example
  - 4.2 Continuous-time Markov chain



A simple two-state Markov chain



Markov process example

- 4.2.1 Infinitesimal definition
  - 4.2.2 Jump chain/holding time definition
  - 4.2.3 Transition probability definition
- 5 Transient evolution
- 6 Properties
  - 6.1 Reducibility
  - 6.2 Periodicity
  - 6.3 Transience and recurrence
    - 6.3.1 Mean recurrence time
    - 6.3.2 Expected number of visits
    - 6.3.3 Absorbing states
  - 6.4 Ergodicity
  - 6.5 Steady-state analysis and limiting distributions
    - 6.5.1 Steady-state analysis and the time-inhomogeneous Markov chain
- 7 Finite state space
  - 7.1 Stationary distribution relation to eigenvectors and simplices
  - 7.2 Time-homogeneous Markov chain with a finite state space
  - 7.3 Convergence speed to the stationary distribution
- 8 Reversible Markov chain
  - 8.1 Closest reversible Markov chain
- 9 Bernoulli scheme
- 10 General state space
  - 10.1 Harris chains
  - 10.2 Locally interacting Markov chains
- 11 Markovian representations
- 12 Transient behaviour
- 13 Stationary distribution
  - 13.1 Example
- 14 Hitting times
  - 14.1 Expected hitting times
- 15 Time reversal
- 16 Embedded Markov chain
- 17 Applications
  - 17.1 Physics
  - 17.2 Chemistry
  - 17.3 Testing
  - 17.4 Speech recognition
  - 17.5 Information sciences
  - 17.6 Queueing theory

- 17.7 Internet applications
- 17.8 Statistics
- 17.9 Economics and finance
- 17.10 Social sciences
- 17.11 Mathematical biology
- 17.12 Genetics
- 17.13 Games
- 17.14 Music
- 17.15 Baseball
- 17.16 Markov text generators
- 17.17 Bioinformatics
- 18 Fitting
- 19 History
- 20 See also
- 21 Notes
- 22 References
- 23 External links

## Introduction

A Markov chain is a stochastic process with the Markov property. The term "Markov chain" refers to the sequence of random variables such a process moves through, with the Markov property defining serial dependence only between adjacent periods (as in a "chain"). It can thus be used for describing systems that follow a chain of linked events, where what happens next depends only on the current state of the system.

The system's state space and time parameter index needs to be specified. The following table gives an overview of the different instances of Markov processes for different levels of state space generality and for discrete time vs. continuous time:

	Countable state space	Continuous or general state space
Discrete-time	(discrete-time) Markov chain on a countable or finite state space	Harris chain (Markov chain on a general state space)
Continuous-time	Continuous-time Markov process	Any continuous stochastic process with the Markov property, e.g. the Wiener process

Note that there is no definitive agreement in the literature on the use of some of the terms that signify special cases of Markov processes. Usually the term "Markov chain" is reserved for a process with a discrete set of times, i.e. a discrete-time Markov chain (DTMC),<sup>[3][3]</sup> but a few authors use the term "Markov process" to refer to a continuous-time Markov chain (CTMC) without explicit mention.<sup>[4][5][6]</sup> In addition, there are other extensions of Markov processes that are referred to as such but do not necessarily fall within any of these four categories (see Markov model). Moreover, the time index need not necessarily be real-valued; like with the state space, there are conceivable processes that move through index sets with other mathematical constructs. Notice that the general state space continuous-time Markov chain is general to such a degree that it has no designated term.

While the time parameter is usually discrete, the state space of a Markov chain does not have any generally agreed-on restrictions: the term may refer to a process on an arbitrary state space.<sup>[7]</sup> However, many applications of Markov chains employ finite or countably infinite (that is, discrete) state spaces, which have a more straightforward statistical analysis. Besides time-index and state-space parameters, there are many other variations, extensions and generalisations (see Variations). For simplicity, most of this article concentrates on the discrete-time, discrete state-space case, unless mentioned otherwise.

The changes of state of the system are called transitions. The probabilities associated with various state changes are called transition probabilities. The process is characterized by a state space, a transition matrix describing the probabilities of particular transitions, and an initial state (or initial distribution) across the state space. By convention, we assume all possible states and transitions have been included in the definition of the process, so there is always a next state, and the process does not terminate.

A discrete-time random process involves a system which is in a certain state at each step, with the state changing randomly between steps. The steps are often thought of as moments in time, but they can equally well refer to physical distance or any other discrete measurement. Formally, the steps are the integers or natural numbers, and the random process is a mapping of these to states. The Markov property states that the conditional probability distribution for the system at the next step (and in fact at all future steps) depends only on the current state of the system, and not additionally on the state of the system at previous steps.

Since the system changes randomly, it is generally impossible to predict with certainty the state of a Markov chain at a given point in the future. However, the statistical properties of the system's future can be predicted. In many applications, it is these statistical properties that are important.

A famous Markov chain is the so-called "drunkard's walk", a random walk on the number line where, at each step, the position may change by +1 or −1 with equal probability. From any position there are two possible transitions, to the next or previous integer. The transition probabilities depend only on the current position, not on the manner in which the position was reached. For example, the transition probabilities from 5 to 4 and 5 to 6 are both 0.5, and all other transition probabilities from 5 are 0. These probabilities are independent of whether the system was previously in 4 or 6.

Another example is the dietary habits of a creature who eats only grapes, cheese, or lettuce, and whose dietary habits conform to the following rules:

- It eats exactly once a day.
- If it ate cheese today, tomorrow it will eat lettuce or grapes with equal probability.



Russian mathematician Andrey Markov.

- If it ate grapes today, tomorrow it will eat grapes with probability 1/10, cheese with probability 4/10 and lettuce with probability 5/10.
- If it ate lettuce today, tomorrow it will eat grapes with probability 4/10 or cheese with probability 6/10. It will not eat lettuce again tomorrow.

This creature's eating habits can be modeled with a Markov chain since its choice tomorrow depends solely on what it ate today, not what it ate yesterday or any other time in the past. One statistical property that could be calculated is the expected percentage, over a long period, of the days on which the creature will eat grapes.

A series of independent events (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.

Many other examples of Markov chains exist.

## Examples

### Gambling

Suppose that you start with \$10, and you wager \$1 on an unending, fair, coin toss indefinitely, or until you lose all of your money. If  $X_n$  represents the number of dollars you have after  $n$  tosses, with  $X_0 = 10$ , then the sequence  $\{X_n : n \in \mathbb{N}\}$  is a Markov process. If I know that you have \$12 now, then it would be expected that with even odds, you will either have \$11 or \$13 after the next toss. This guess is not improved by the added knowledge that you started with \$10, then went up to \$11, down to \$10, up to \$11, and then to \$12.

The process described here is a Markov chain on a countable state space that follows a random walk.

### A birth-death process

If one pops one hundred kernels of popcorn, each kernel popping at an independent exponentially-distributed time, then this would be a continuous-time Markov process. If  $X_t$  denotes the number of kernels which have popped up to time  $t$ , the problem can be defined as finding the number of kernels that will pop in some later time. The only thing one needs to know is the number of kernels that have popped prior to the time " $t$ ". It is not necessary to know *when* they popped, so knowing  $X_t$  for previous times " $t$ " is not relevant.

The process described here is an approximation of a Poisson point process - Poisson processes are also Markov processes.

### A non-Markov example

Suppose that you have a coin purse containing five quarters (each worth 25c), five nickels (each worth 5c) and five dimes (each worth 10c), and one-by-one, you randomly draw coins from the purse and set them on a table. If  $X_n$  represents the total value of the coins set on the table after  $n$  draws, with  $X_0 = 0$ , then the sequence  $\{X_n : n \in \mathbb{N}\}$  is *not* a Markov process.

To see why this is the case, suppose that in your first six draws, you draw all five nickels, and then a quarter. So  $\mathbf{X}_6 = \$0.50$ . If we know not just  $\mathbf{X}_6$ , but the earlier values as well, then we can determine which coins have been drawn, and we know that the next coin will not be a nickel, so we can determine that  $\mathbf{X}_7 \geq \$0.60$  with probability 1. But if we do not know the earlier values, then based only on the value  $\mathbf{X}_6$  we might guess that we had drawn four dimes and two nickels, in which case it would certainly be possible to draw another nickel next. Thus, our guesses about  $\mathbf{X}_7$  are impacted by our knowledge of values prior to  $\mathbf{X}_6$ .

## Markov property

### The general case

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space with a filtration  $(\mathcal{F}_t, t \in T)$ , for some (totally ordered) index set  $T$ ; and let  $(S, \mathcal{S})$  be a measure space. An  $S$ -valued stochastic process  $\mathbf{X} = (X_t, t \in T)$  adapted to the filtration is said to possess the **Markov property** with respect to the  $\{\mathcal{F}_t\}$  if, for each  $A \in \mathcal{S}$  and each  $s, t \in T$  with  $s < t$ ,

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s).^{[8]}$$

A **Markov process** is a stochastic process which satisfies the Markov property with respect to its natural filtration.

### For discrete-time Markov chains

In the case where  $S$  is a discrete set with the discrete sigma algebra and  $T = \mathbb{N}$ , this can be reformulated as follows:

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}).$$

## Formal definition

### Discrete-time Markov chain

A discrete-time Markov chain is a sequence of random variables  $X_1, X_2, X_3, \dots$  with the Markov property, namely that the probability of moving to next state depends only on the present state and not on the previous states

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n), \text{ if both conditional probabilities are well defined, i.e. if } \Pr(X_1 = x_1, \dots, X_n = x_n) > 0.$$

The possible values of  $X_i$  form a countable set  $S$  called the **state space** of the chain.

Markov chains are often described by a sequence of directed graphs, where the edges of graph  $n$  are labeled by the probabilities of going from one state at time  $n$  to the other states at time  $n+1$ ,  $\Pr(X_{n+1} = x \mid X_n = x_n)$ . The same information is represented by the transition matrix from time  $n$  to time  $n+1$ . However, Markov chains are frequently assumed to be time-homogeneous (see variations below), in which case the graph and matrix are independent of  $n$  and are thus not presented as sequences.

These descriptions highlight the structure of the Markov chain that is independent of the initial distribution  $\Pr(X_1 = x_1)$ . When time-homogeneous, the chain can be interpreted as a state machine assigning a probability of hopping from each vertex or state to an adjacent one. The probability  $\Pr(X_n = x \mid X_1 = x_1)$  of the machine's state can be analyzed as the statistical behavior of the machine with an element  $x_1$  of the state space as input, or as the behavior of the machine with the initial distribution  $\Pr(X_1 = y) = [x_1 = y]$  of states as input, where  $[P]$  is the Iverson bracket.

The fact that some sequences of states might have zero probability of occurring corresponds to a graph with multiple connected components, where we omit edges that would carry a zero transition probability. For example, if  $a$  has a nonzero probability of going to  $b$ , but  $a$  and  $x$  lie in different connected components of the graph, then  $\Pr(X_{n+1} = b \mid X_n = a)$  is defined, while  $\Pr(X_{n+1} = b \mid X_1 = x, \dots, X_n = a)$  is not.

## Variations

- **Time-homogeneous Markov chains** (or **stationary Markov chains**) are processes where

$$\Pr(X_{n+1} = x \mid X_n = y) = \Pr(X_n = x \mid X_{n-1} = y)$$

for all  $n$ . The probability of the transition is independent of  $n$ .

- A **Markov chain of order  $m$**  (or a Markov chain with memory  $m$ ), where  $m$  is finite, is a process satisfying

$$\begin{aligned} &\Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_1 = x_1) \\ &= \Pr(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_{n-m} = x_{n-m}) \text{ for } n > m \end{aligned}$$

In other words, the future state depends on the past  $m$  states. It is possible to construct a chain  $(Y_n)$  from  $(X_n)$  which has the 'classical' Markov property by taking as state space the ordered  $m$ -tuples of  $X$  values, ie.  $Y_n = (X_n, X_{n-1}, \dots, X_{n-m+1})$ .

## Example

A state diagram for a simple example is shown in the figure on the right, using a directed graph to picture the state transitions. The states represent whether a hypothetical stock market is exhibiting a bull market, bear market, or stagnant market trend during a given week. According to the figure, a bull week is followed by another bull week 90% of the time, a bear week 7.5% of the time, and a stagnant week the other 2.5% of the time. Labelling the state space  $\{1 = \text{bull}, 2 = \text{bear}, 3 = \text{stagnant}\}$  the transition matrix for this example is

$$P = \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}.$$

The distribution over states can be written as a stochastic row vector  $x$  with the relation  $x^{(n+1)} = x^{(n)}P$ . So if at time  $n$  the system is in state  $x^{(n)}$ , then three time periods later, at time  $n+3$  the distribution is

$$\begin{aligned} x^{(n+3)} &= x^{(n+2)}P = (x^{(n+1)}P)P \\ &= x^{(n+1)}P^2 = (x^{(n)}P)P^2 \\ &= x^{(n)}P^3 \end{aligned}$$

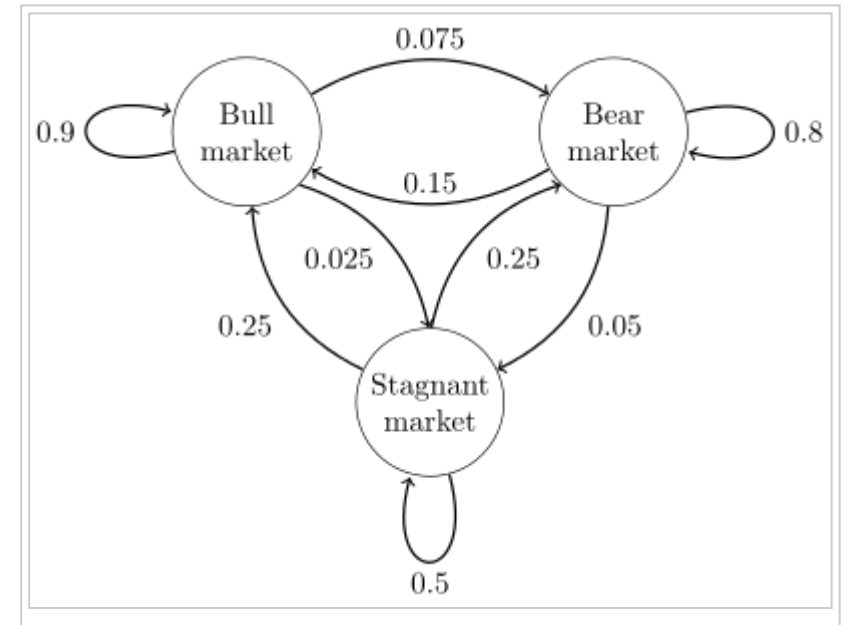
In particular, if at time  $n$  the system is in state 2 (bear), then at time  $n+3$  the distribution is

$$\begin{aligned} x^{(n+3)} &= [0 \quad 1 \quad 0] \begin{bmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}^3 \\ &= [0 \quad 1 \quad 0] \begin{bmatrix} 0.7745 & 0.17875 & 0.04675 \\ 0.3575 & 0.56825 & 0.07425 \\ 0.4675 & 0.37125 & 0.16125 \end{bmatrix} \\ &= [0.3575 \quad 0.56825 \quad 0.07425]. \end{aligned}$$

Using the transition matrix it is possible to calculate, for example, the long-term fraction of weeks during which the market is stagnant, or the average number of weeks it will take to go from a stagnant to a bull market. Using the transition probabilities, the steady-state probabilities indicate that 62.5% of weeks will be in a bull market, 31.25% of weeks will be in a bear market and 6.25% of weeks will be stagnant, since:

$$\lim_{N \rightarrow \infty} P^N = \begin{bmatrix} 0.625 & 0.3125 & 0.0625 \\ 0.625 & 0.3125 & 0.0625 \\ 0.625 & 0.3125 & 0.0625 \end{bmatrix}$$

A thorough development and many examples can be found in the on-line monograph Meyn & Tweedie 2005.<sup>[9]</sup>





A finite state machine can be used as a representation of a Markov chain. Assuming a sequence of independent and identically distributed input signals (for example, symbols from a binary alphabet chosen by coin tosses), if the machine is in state  $y$  at time  $n$ , then the probability that it moves to state  $x$  at time  $n + 1$  depends only on the current state.

## Continuous-time Markov chain

A continuous-time Markov chain  $(X_t)_{t \geq 0}$  is defined by a finite or countable state space  $S$ , a transition rate matrix  $Q$  with dimensions equal to that of the state space and initial probability distribution defined on the state space. For  $i \neq j$ , the elements  $q_{ij}$  are non-negative and describe the rate of the process transitions from state  $i$  to state  $j$ . The elements  $q_{ii}$  are chosen such that each row of the transition rate matrix sums to zero.

There are three equivalent definitions of the process.<sup>[10]</sup>

### Infinitesimal definition

Let  $X_t$  be the random variable describing the state of the process at time  $t$ , and assume that the process is in a state  $i$  at time  $t$ . Then  $X_{t+h}$  is independent of previous values  $(X_s : s \leq t)$  and as  $h \rightarrow 0$  uniformly in  $t$  for all  $j$

$$\Pr(X(t+h) = j | X(t) = i) = \delta_{ij} + q_{ij}h + o(h)$$

using little-o notation. The  $q_{ij}$  can be seen as measuring how quickly the transition from  $i$  to  $j$  happens

### Jump chain/holding time definition

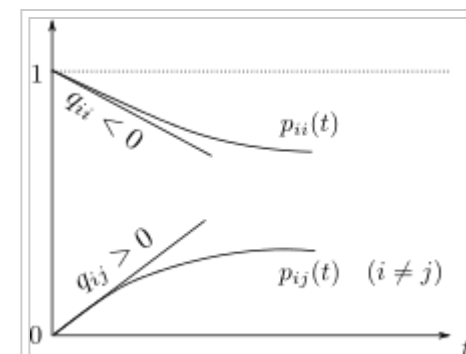
Define a discrete-time Markov chain  $Y_n$  to describe the  $n$ th jump of the process and variables  $S_1, S_2, S_3, \dots$  to describe holding times in each of the states where  $S_i$  follows the exponential distribution with rate parameter  $-q_{Y_i Y_i}$ .

### Transition probability definition

For any value  $n = 0, 1, 2, 3, \dots$  and times indexed up to this value of  $n$ :  $t_0, t_1, t_2, \dots$  and all states recorded at these times  $i_0, i_1, i_2, i_3, \dots$  it holds that

$$\Pr(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, X_{t_1} = i_1, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$$

where  $p_{ij}$  is the solution of the forward equation (a first-order differential equation)



The continuous time Markov chain is characterized by the transition rates, the derivatives with respect to time of the transition probabilities between states  $i$  and  $j$ .

$$P'(t) = P(t)Q$$

with initial condition  $P(0)$  is the identity matrix.

## Transient evolution

The probability of going from state  $i$  to state  $j$  in  $n$  time steps is

$$p_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i)$$

and the single-step transition is

$$p_{ij} = \Pr(X_1 = j \mid X_0 = i).$$

For a time-homogeneous Markov chain:

$$p_{ij}^{(n)} = \Pr(X_{k+n} = j \mid X_k = i)$$

and

$$p_{ij} = \Pr(X_{k+1} = j \mid X_k = i).$$

The  $n$ -step transition probabilities satisfy the Chapman–Kolmogorov equation, that for any  $k$  such that  $0 < k < n$ ,

$$p_{ij}^{(n)} = \sum_{r \in S} p_{ir}^{(k)} p_{rj}^{(n-k)}$$

where  $S$  is the state space of the Markov chain.

The marginal distribution  $\Pr(X_n = x)$  is the distribution over states at time  $n$ . The initial distribution is  $\Pr(X_0 = x)$ . The evolution of the process through one time step is described by

$$\Pr(X_n = j) = \sum_{r \in S} p_{rj} \Pr(X_{n-1} = r) = \sum_{r \in S} p_{rj}^{(n)} \Pr(X_0 = r).$$

Note: The superscript  $(n)$  is an index and not an exponent.

# Properties

## Reducibility

A state  $j$  is said to be **accessible** from a state  $i$  (written  $i \rightarrow j$ ) if a system started in state  $i$  has a non-zero probability of transitioning into state  $j$  at some point. Formally, state  $j$  is accessible from state  $i$  if there exists an integer  $n_{ij} \geq 0$  such that

$$\Pr(X_{n_{ij}} = j \mid X_0 = i) = p_{ij}^{(n_{ij})} > 0.$$

This integer is allowed to be different for each pair of states, hence the subscripts in  $n_{ij}$ . Allowing  $n$  to be zero means that every state is defined to be accessible from itself.

A state  $i$  is said to **communicate** with state  $j$  (written  $i \leftrightarrow j$ ) if both  $i \rightarrow j$  and  $j \rightarrow i$ . A set of states  $C$  is a **communicating class** if every pair of states in  $C$  communicates with each other. A communicating class is **closed** if the probability of leaving the class is zero, namely that if  $i$  is in  $C$  but  $j$  is not, then  $j$  is not accessible from  $i$ . It can be shown that communication in this sense is an equivalence relation and thus that communicating classes are the equivalence classes of this relation.

The set of communicating classes forms a directed, acyclic graph by inheriting the arrows from the original state space. A communicating class is closed if and only if it has no outgoing arrows in this graph.

A state  $i$  is said to be **essential** or **final** if for all  $j$  such that  $i \rightarrow j$  it is also true that  $j \rightarrow i$ . A state  $i$  is **inessential** if it is not essential.<sup>[11]</sup> A state is final if and only if its communicating class is closed.

A Markov chain is said to be **irreducible** if its state space is a single communicating class; in other words, if it is possible to get to any state from any state.

## Periodicity

A state  $i$  has **period**  $k$  if any return to state  $i$  must occur in multiples of  $k$  time steps. Formally, the period of a state is defined as

$$k = \gcd\{n > 0 : \Pr(X_n = i \mid X_0 = i) > 0\}$$

(where "gcd" is the greatest common divisor) provided that this set is not empty. Otherwise the period is not defined. Note that even though a state has period  $k$ , it may not be possible to reach the state in  $k$  steps. For example, suppose it is possible to return to the state in  $\{6, 8, 10, 12, \dots\}$  time steps;  $k$  would be 2, even though 2 does not appear in this list.

If  $k = 1$ , then the state is said to be **aperiodic**: returns to state  $i$  can occur at irregular times. It can be demonstrated that a state  $i$  is **aperiodic** if and only if there exists  $n$  such that for all  $n' \geq n$ ,

$$\Pr(X_{n'} = i \mid X_0 = i) > 0.$$

Otherwise ( $k > 1$ ), the state is said to be **periodic with period  $k$** . A Markov chain is aperiodic if every state is aperiodic. An irreducible Markov chain only needs one aperiodic state to imply all states are aperiodic.

Every state of a bipartite graph has an even **period**.

## Transience and recurrence

A state  $i$  is said to be **transient** if, given that we start in state  $i$ , there is a non-zero probability that we will never return to  $i$ . Formally, let the random variable  $T_i$  be the first return time to state  $i$  (the "**hitting time**"):

$$T_i = \inf\{n \geq 1 : X_n = i \mid X_0 = i\}.$$

The number

$$f_{ii}^{(n)} = \Pr(T_i = n)$$

is the probability that we return to state  $i$  for the first time after  $n$  steps. Therefore, state  $i$  is transient if

$$\Pr(T_i < \infty) = \sum_{n=1}^{\infty} f_{ii}^{(n)} < 1.$$

State  $i$  is **recurrent** (or **persistent**) if it is not transient. Recurrent states are guaranteed (with probability 1) to have a finite hitting time. Recurrence and transience are class properties, that is, they either hold or do not hold equally for all members of a communicating class.

## Mean recurrence time

Even if the hitting time is finite with probability 1, it need not have a finite expectation. The **mean recurrence time** at state  $i$  is the expected return time  $M_i$ :

$$M_i = E[T_i] = \sum_{n=1}^{\infty} n \cdot f_{ii}^{(n)}.$$

State  $i$  is **positive recurrent** (or **non-null persistent**) if  $M_i$  is finite; otherwise, state  $i$  is **null recurrent** (or **null persistent**).

### Expected number of visits

It can be shown that a state  $i$  is recurrent if and only if the expected number of visits to this state is infinite, i.e.,

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.$$

### Absorbing states

A state  $i$  is called **absorbing** if it is impossible to leave this state. Therefore, the state  $i$  is absorbing if and only if

$$p_{ii} = 1 \text{ and } p_{ij} = 0 \text{ for } i \neq j.$$

If every state can reach an absorbing state, then the Markov chain is an absorbing Markov chain.

### Ergodicity

A state  $i$  is said to be **ergodic** if it is aperiodic and positive recurrent. In other words, a state  $i$  is ergodic if it is recurrent, has a period of 1 and it has finite mean recurrence time. If all states in an irreducible Markov chain are ergodic, then the chain is said to be ergodic.

It can be shown that a finite state irreducible Markov chain is ergodic if it has an aperiodic state. A Markov chain has the ergodic property if there's a finite number  $N$  such that any state can be reached from any other state in exactly  $N$  steps. In case of a fully connected transition matrix where all transitions have a non-zero probability, this condition is fulfilled with  $N=1$ . A Markov chain with more than one state and just one out-going transition per state cannot be ergodic.

### Steady-state analysis and limiting distributions

If the Markov chain is a time-homogeneous Markov chain, so that the process is described by a single, time-independent matrix  $p_{ij}$ , then the vector  $\pi$  is called a **stationary distribution** (or **invariant measure**) if  $\forall j \in S$  it satisfies

$$0 \leq \pi_j \leq 1.$$

$$\sum_{j \in S} \pi_j = 1.$$

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}.$$

An irreducible chain has a stationary distribution if and only if all of its states are positive recurrent.<sup>[12]</sup> In that case,  $\pi$  is unique and is related to the expected return time:

$$\pi_j = \frac{C}{M_j},$$

where  $C$  is the normalizing constant. Further, if the positive recurrent chain is both irreducible and aperiodic, it is said to have a *limiting* distribution; for any  $i$  and  $j$ ,

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{C}{M_j}.$$

Note that there is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it begins. Such  $\pi$  is called the **equilibrium distribution** of the chain.

If a chain has more than one closed communicating class, its stationary distributions will not be unique (consider any closed communicating class  $C_i$  in the chain; each one will have its own unique stationary distribution  $\pi_i$ . Extending these distributions to the overall chain, setting all values to zero outside the communication class, yields that the set of invariant measures of the original chain is the set of all convex combinations of the  $\pi_i$ 's). However, if a state  $j$  is aperiodic, then

$$\lim_{n \rightarrow \infty} p_{jj}^{(n)} = \frac{C}{M_j}$$

and for any other state  $i$ , let  $f_{ij}$  be the probability that the chain ever visits state  $j$  if it starts at  $i$ ,

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = C \frac{f_{ij}}{M_j}.$$

If a state  $i$  is periodic with period  $k > 1$  then the limit

$$\lim_{n \rightarrow \infty} p_{ii}^{(n)}$$

does not exist, although the limit

$$\lim_{n \rightarrow \infty} p_{ii}^{(kn+r)}$$

does exist for every integer  $r$ .

### Steady-state analysis and the time-inhomogeneous Markov chain

A Markov chain need not necessarily be time-homogeneous to have an equilibrium distribution. If there is a probability distribution over states  $\boldsymbol{\pi}$  such that

$$\pi_j = \sum_{i \in S} \pi_i \Pr(X_{n+1} = j \mid X_n = i)$$

for every state  $j$  and every time  $n$  then  $\boldsymbol{\pi}$  is an equilibrium distribution of the Markov chain. Such can occur in Markov chain Monte Carlo (MCMC) methods in situations where a number of different transition matrices are used, because each is efficient for a particular kind of mixing, but each matrix respects a shared equilibrium distribution.

## Finite state space

If the state space is finite, the transition probability distribution can be represented by a matrix, called the **transition matrix**, with the  $(i, j)$ th element of  $\mathbf{P}$  equal to

$$p_{ij} = \Pr(X_{n+1} = j \mid X_n = i).$$

Since each row of  $\mathbf{P}$  sums to one and all elements are non-negative,  $\mathbf{P}$  is a right stochastic matrix.

### Stationary distribution relation to eigenvectors and simplices

A stationary distribution  $\boldsymbol{\pi}$  is a (row) vector, whose entries are non-negative and sum to 1, is unchanged by the operation of transition matrix  $\mathbf{P}$  on it and so is defined by

$$\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}.$$

By comparing this definition with that of an eigenvector we see that the two concepts are related and that

$$\boldsymbol{\pi} = \frac{\mathbf{e}}{\sum_i e_i}$$

is a normalized ( $\sum_i \pi_i = 1$ ) multiple of a left eigenvector  $\mathbf{e}$  of the transition matrix  $\mathbf{P}^T$  with an eigenvalue of 1. If there is more than one unit eigenvector then a weighted sum of the corresponding stationary states is also a stationary state. But for a Markov chain one is usually more interested in a stationary state that is the limit of the sequence of distributions for some initial distribution.

The values of a stationary distribution  $\pi_i$  are associated with the state space of  $\mathbf{P}$  and its eigenvectors have their relative proportions preserved. Since the components of  $\pi$  are positive and the constraint that their sum is unity can be rewritten as  $\sum_i \mathbf{1} \cdot \pi_i = 1$  we see that the dot product of  $\pi$  with a vector whose components are all 1 is unity and that  $\pi$  lies on a simplex.

## Time-homogeneous Markov chain with a finite state space

If the Markov chain is time-homogeneous, then the transition matrix  $\mathbf{P}$  is the same after each step, so the  $k$ -step transition probability can be computed as the  $k$ -th power of the transition matrix,  $\mathbf{P}^k$ .

If the Markov chain is irreducible and aperiodic, then there is a unique stationary distribution  $\pi$ . Additionally, in this case  $\mathbf{P}^k$  converges to a rank-one matrix in which each row is the stationary distribution  $\pi$ , that is,

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \mathbf{1}\pi$$

where  $\mathbf{1}$  is the column vector with all entries equal to 1. This is stated by the Perron–Frobenius theorem. If, by whatever means,  $\lim_{k \rightarrow \infty} \mathbf{P}^k$  is found, then the stationary distribution of the Markov chain in question can be easily determined for any starting distribution, as will be explained below.

For some stochastic matrices  $\mathbf{P}$ , the limit  $\lim_{k \rightarrow \infty} \mathbf{P}^k$  does not exist while the stationary distribution does, as shown by this example:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \mathbf{P}^{2k} = I \quad \mathbf{P}^{2k+1} = \mathbf{P}$$

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

Note that this example illustrates a periodic Markov chain.

Because there are a number of different special cases to consider, the process of finding this limit if it exists can be a lengthy task. However, there are many techniques that can assist in finding this limit. Let  $\mathbf{P}$  be an  $n \times n$  matrix, and define  $\mathbf{Q} = \lim_{k \rightarrow \infty} \mathbf{P}^k$ .

It is always true that



$$\mathbf{Q}\mathbf{P} = \mathbf{Q}.$$

Subtracting  $\mathbf{Q}$  from both sides and factoring then yields

$$\mathbf{Q}(\mathbf{P} - \mathbf{I}_n) = \mathbf{0}_{n,n},$$

where  $\mathbf{I}_n$  is the identity matrix of size  $n$ , and  $\mathbf{0}_{n,n}$  is the zero matrix of size  $n \times n$ . Multiplying together stochastic matrices always yields another stochastic matrix, so  $\mathbf{Q}$  must be a stochastic matrix (see the definition above). It is sometimes sufficient to use the matrix equation above and the fact that  $\mathbf{Q}$  is a stochastic matrix to solve for  $\mathbf{Q}$ . Including the fact that the sum of each the rows in  $\mathbf{P}$  is 1, there are  $n+1$  equations for determining  $n$  unknowns, so it is computationally easier if on the one hand one selects one row in  $\mathbf{Q}$  and substitute each of its elements by one, and on the other one substitute the corresponding element (the one in the same column) in the vector  $\mathbf{0}$ , and next left-multiply this latter vector by the inverse of transformed former matrix to find  $\mathbf{Q}$ .

Here is one method for doing so: first, define the function  $f(\mathbf{A})$  to return the matrix  $\mathbf{A}$  with its right-most column replaced with all 1's. If  $[f(\mathbf{P} - \mathbf{I}_n)]^{-1}$  exists then

$$\mathbf{Q} = f(\mathbf{0}_{n,n})[f(\mathbf{P} - \mathbf{I}_n)]^{-1}.$$

Explain: The original matrix equation is equivalent to a system of  $n \times n$  linear equations in  $n \times n$  variables. And there are  $n$  more linear equations from the fact that  $\mathbf{Q}$  is a right stochastic matrix whose each row sums to 1. So it needs any  $n \times n$  independent linear equations of the  $(n \times n + n)$  equations to solve for the  $n \times n$  variables. In this example, the  $n$  equations from “ $\mathbf{Q}$  multiplied by the right-most column of  $(\mathbf{P} - \mathbf{I}_n)$ ” have been replaced by the  $n$  stochastic ones.

One thing to notice is that if  $\mathbf{P}$  has an element  $\mathbf{P}_{i,i}$  on its main diagonal that is equal to 1 and the  $i$ th row or column is otherwise filled with 0's, then that row or column will remain unchanged in all of the subsequent powers  $\mathbf{P}^k$ . Hence, the  $i$ th row or column of  $\mathbf{Q}$  will have the 1 and the 0's in the same positions as in  $\mathbf{P}$ .

## Convergence speed to the stationary distribution

As stated earlier, from the equation  $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ , (if exists) the stationary (or steady state) distribution  $\boldsymbol{\pi}$  is a left eigenvector of row stochastic matrix  $\mathbf{P}$ . Then assuming that  $\mathbf{P}$  is diagonalizable or equivalently that  $\mathbf{P}$  has  $n$  linearly independent eigenvectors, speed of convergence is elaborated as follows. (For non-diagonalizable, i.e. defective matrices, one may start with the Jordan normal form of  $\mathbf{P}$  and proceed with a bit more involved set of arguments in a similar way.<sup>[13]</sup>)

Let  $\mathbf{U}$  be the matrix of eigenvectors (each normalized to having an L2 norm equal to 1) where each column is a left eigenvector of  $\mathbf{P}$  and let  $\boldsymbol{\Sigma}$  be the diagonal matrix of left eigenvalues of  $\mathbf{P}$ , i.e.  $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$ . Then by eigendecomposition

$$\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{-1}.$$

Let the eigenvalues be enumerated such that  $1 = |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ . Since  $\mathbf{P}$  is a row stochastic matrix, its largest left eigenvalue is 1. If there is a unique stationary distribution, then the largest eigenvalue and the corresponding eigenvector is unique too (because there is no other  $\boldsymbol{\pi}$  which solves the stationary distribution equation above). Let  $\mathbf{u}_i$  be the  $i$ th column of  $\mathbf{U}$  matrix, i.e.  $\mathbf{u}_i$  is the left eigenvector of  $\mathbf{P}$  corresponding to  $\lambda_i$ . Also let  $\mathbf{x}$  be a length  $n$  row vector that

represents a valid probability distribution; since the eigenvectors  $\mathbf{u}_i$  span  $\mathbf{R}^n$ , we can write

$$\mathbf{x}^T = \sum_{i=1}^n a_i \mathbf{u}_i$$

for some set of  $a_i \in \mathbb{R}$ . If we start multiplying  $\mathbf{P}$  with  $\mathbf{x}$  from left and continue this operation with the results, in the end we get the stationary distribution  $\boldsymbol{\pi}$ . In other words,  $\boldsymbol{\pi} = \mathbf{u}_i \leftarrow \mathbf{x} \mathbf{P} \mathbf{P} \dots \mathbf{P} = \mathbf{x} \mathbf{P}^k$  as  $k$  goes to infinity. That means

$$\begin{aligned} \boldsymbol{\pi}^{(k)} &= \mathbf{x} (\mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^{-1}) (\mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^{-1}) \dots (\mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^{-1}) \\ &= \mathbf{x} \mathbf{U} \boldsymbol{\Sigma}^k \mathbf{U}^{-1} \end{aligned}$$

since  $\mathbf{U} \mathbf{U}^{-1} = \mathbf{I}$  the identity matrix and power of a diagonal matrix is also a diagonal matrix where each entry is taken to that power.

$$\begin{aligned} &= (a_1 \mathbf{u}_1^T + a_2 \mathbf{u}_2^T + \dots + a_n \mathbf{u}_n^T) \mathbf{U} \boldsymbol{\Sigma}^k \mathbf{U}^{-1}, \\ &= a_1 \lambda_1^k \mathbf{u}_1 + a_2 \lambda_2^k \mathbf{u}_2 + \dots + a_n \lambda_n^k \mathbf{u}_n, \end{aligned}$$

since the eigenvectors are orthonormal. Then<sup>[14]</sup>

$$= \lambda_1^k \left\{ a_1 \mathbf{u}_1 + a_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \mathbf{u}_2 + a_3 \left( \frac{\lambda_3}{\lambda_1} \right)^k \mathbf{u}_3 + \dots + a_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \mathbf{u}_n \right\}.$$

Since  $\boldsymbol{\pi} = \mathbf{u}_1$ ,  $\boldsymbol{\pi}^{(k)}$  approaches to  $\boldsymbol{\pi}$  as  $k$  goes to infinity with a speed in the order of  $\lambda_2/\lambda_1$  exponentially. This follows because  $|\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$ , hence  $\lambda_2/\lambda_1$  is the dominant term. Random noise in the state distribution  $\boldsymbol{\pi}$  can also speed up this convergence to the stationary distribution.<sup>[15]</sup>

## Reversible Markov chain

A Markov chain is said to be **reversible** if there is a probability distribution  $\boldsymbol{\pi}$  over its states such that

$$\pi_i \Pr(X_{n+1} = j \mid X_n = i) = \pi_j \Pr(X_{n+1} = i \mid X_n = j)$$

for all times  $n$  and all states  $i$  and  $j$ . This condition is known as the **detailed balance** condition (some books call it the local balance equation).

Considering a fixed arbitrary time  $n$  and using the shorthand

$$p_{ij} = \Pr(X_{n+1} = j \mid X_n = i),$$

the detailed balance equation can be written more compactly as

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

The single time-step from  $n$  to  $n+1$  can be thought of as each person  $i$  having  $\pi_i$  dollars initially and paying each person  $j$  a fraction  $p_{ij}$  of it. The detailed balance condition states that upon each payment, the other person pays exactly the same amount of money back.<sup>[16]</sup> Clearly the total amount of money  $\pi$  each person has remains the same after the time-step, since every dollar spent is balanced by a corresponding dollar received. This can be shown more formally by the equality

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j,$$

which essentially states that the total amount of money person  $j$  receives (including from himself) during the time-step equals the amount of money he pays others, which equals all the money he initially had because it was assumed that all money is spent (i.e.  $p_{ji}$  sums to 1 over  $i$ ). The assumption is a technical one, because the money not really used is simply thought of as being paid from person  $j$  to himself (i.e.  $p_{jj}$  is not necessarily zero).

As  $n$  was arbitrary, this reasoning holds for any  $n$ , and therefore for reversible Markov chains  $\pi$  is always a steady-state distribution of  $\Pr(X_{n+1} = j \mid X_n = i)$  for every  $n$ .

If the Markov chain begins in the steady-state distribution, *i.e.*, if  $\Pr(X_0 = i) = \pi_i$ , then  $\Pr(X_n = i) = \pi_i$  for all  $n$  and the detailed balance equation can be written as

$$\Pr(X_n = i, X_{n+1} = j) = \Pr(X_{n+1} = i, X_n = j).$$

The left- and right-hand sides of this last equation are identical except for a reversing of the time indices  $n$  and  $n + 1$ .

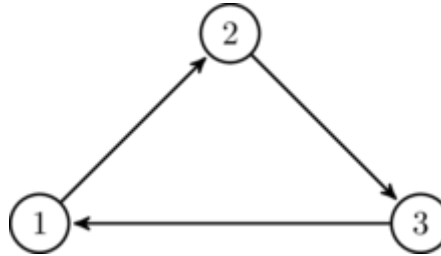
Kolmogorov's criterion gives a necessary and sufficient condition for a Markov chain to be reversible directly from the transition matrix probabilities. The criterion requires that the products of probabilities around every closed loop are the same in both directions around the loop.

Reversible Markov chains are common in Markov chain Monte Carlo (MCMC) approaches because the detailed balance equation for a desired distribution  $\pi$  necessarily implies that the Markov chain has been constructed so that  $\pi$  is a steady-state distribution. Even with time-inhomogeneous Markov chains, where multiple transition matrices are used, if each such transition matrix exhibits detailed balance with the desired  $\pi$  distribution, this necessarily implies that  $\pi$  is a steady-state distribution of the Markov chain.

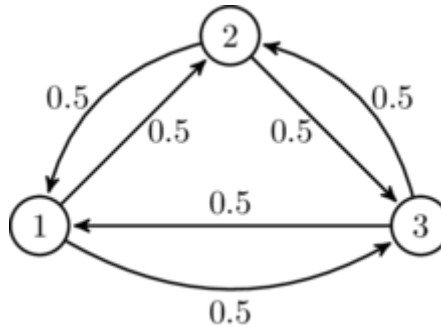
## Closest reversible Markov chain

For any time-homogeneous Markov chain given by a transition matrix  $\mathbf{P} \in \mathbb{R}^{n \times n}$ , any norm  $\|\cdot\|$  on  $\mathbb{R}^{n \times n}$  which is induced by a scalar product, and any probability vector  $\boldsymbol{\pi}$ , there exists a unique transition matrix  $\mathbf{P}^*$  which is reversible according to  $\boldsymbol{\pi}$  and which is closest to  $\mathbf{P}$  according to the norm  $\|\cdot\|$ . The matrix  $\mathbf{P}^*$  can be computed by solving a quadratic-convex optimization problem.<sup>[17]</sup>

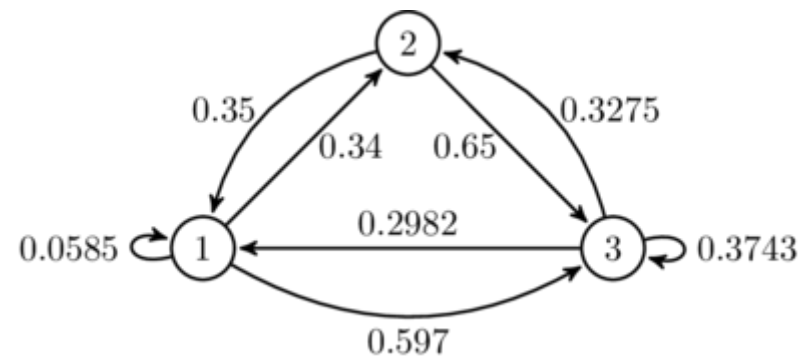
For example, consider the following Markov chain:



This Markov chain is not reversible. According to the Frobenius Norm the closest reversible Markov chain according to  $\boldsymbol{\pi} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$  can be computed as



If we choose the probability vector randomly as  $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$ , then the closest reversible Markov chain according to the Frobenius norm is approximately given by



## Bernoulli scheme

A Bernoulli scheme is a special case of a Markov chain where the transition probability matrix has identical rows, which means that the next state is even independent of the current state (in addition to being independent of the past states). A Bernoulli scheme with only two possible states is known as a Bernoulli process.

## General state space

For an overview of Markov chains on a general state space, see the article Markov chains on a measurable state space.

## Harris chains

Many results for Markov chains with finite state space can be generalized to chains with uncountable state space through Harris chains. The main idea is to see if there is a point in the state space that the chain hits with probability one. Generally, it is not true for continuous state space, however, we can define sets  $A$  and  $B$  along with a positive number  $\varepsilon$  and a probability measure  $\rho$ , such that

$$\begin{cases} \text{If } \tau_A = \inf\{n \geq 0 : X_n \in A\}, & \text{then } P_z(\tau_A < \infty) > 0 \text{ for all } z. \\ \text{If } x \in A \text{ and } C \subset B, & \text{then } p(x, C) \geq \varepsilon \rho(C). \end{cases}$$

Then we could collapse the sets into an auxiliary point  $\alpha$ , and a recurrent Harris chain can be modified to contain  $\alpha$ . Lastly, the collection of Harris chains is a comfortable level of generality, which is broad enough to contain a large number of interesting examples, yet restrictive enough to allow for a rich theory.

The use of Markov chains in Markov chain Monte Carlo methods covers cases where the process follows a continuous state space.

## Locally interacting Markov chains

Considering a collection of Markov chains whose evolution takes in account the state of other Markov chains, is related to the notion of locally interacting Markov chains. This corresponds to the situation when the state space has a (Cartesian-) product form. See interacting particle system and stochastic cellular automata (probabilistic cellular automata). See for instance *Interaction of Markov Processes*<sup>[18]</sup> or<sup>[19]</sup>

## Markovian representations

In some cases, apparently non-Markovian processes may still have Markovian representations, constructed by expanding the concept of the 'current' and 'future' states. For example, let  $X$  be a non-Markovian process. Then define a process  $Y$ , such that each state of  $Y$  represents a time-interval of states of  $X$ . Mathematically, this takes the form:

$$Y(t) = \{X(s) : s \in [a(t), b(t)]\}.$$

If  $Y$  has the Markov property, then it is a Markovian representation of  $X$ .

An example of a non-Markovian process with a Markovian representation is an autoregressive time series of order greater than one.<sup>[20]</sup>

## Transient behaviour

Write  $P(t)$  for the matrix with entries  $p_{ij} = P(X_t = j \mid X_0 = i)$ . Then the matrix  $P(t)$  satisfies the forward equation, a first-order differential equation

$$P'(t) = P(t)Q$$

where the prime denotes differentiation with respect to  $t$ . The solution to this equation is given by a matrix exponential

$$P(t) = e^{tQ}$$

In a simple case such as a CTMC on the state space  $\{1,2\}$ . The general  $Q$  matrix for such a process is the following  $2 \times 2$  matrix with  $\alpha, \beta > 0$

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

The above relation for forward matrix can be solved explicitly in this case to give

$$P(t) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} \end{pmatrix}$$

However, direct solutions are complicated to compute for larger matrices. The fact that  $Q$  is the generator for a semigroup of matrices

$$P(t+s) = e^{(t+s)Q} = e^{tQ}e^{sQ} = P(t)P(s)$$

is used.

## Stationary distribution

The stationary distribution for an irreducible recurrent CTMC is the probability distribution to which the process converges for large values of  $t$ . Observe that for the two-state process considered earlier with  $P(t)$  given by

$$P(t) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t} \\ \frac{\beta}{\alpha+\beta} - \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} & \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t} \end{pmatrix}$$

as  $t \rightarrow \infty$  the distribution tends to

$$P_{\pi} = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$$

Observe that each row has the same distribution as this does not depend on starting state. The row vector  $\pi$  may be found by solving<sup>[21]</sup>

$$\pi Q = 0.$$

with the additional constraint that

$$\sum_{i \in \mathcal{S}} \pi_i = 1.$$

## Example

The image to the right describes a continuous-time Markov chain with state-space {Bull market, Bear market, Stagnant market} and transition rate matrix

$$Q = \begin{pmatrix} -0.025 & 0.02 & 0.005 \\ 0.3 & -0.5 & 0.2 \\ 0.02 & 0.4 & -0.42 \end{pmatrix}.$$

The stationary distribution of this chain can be found by solving  $\pi Q = 0$  subject to the constraint that elements must sum to 1 to obtain

$$\pi = (0.885 \quad 0.071 \quad 0.044).$$

## Hitting times

The hitting time is the time, starting in a given set of states until the chain arrives in a given state or set of states. The distribution of such a time period has a phase type distribution. The simplest such distribution is that of a single exponentially distributed transition.

### Expected hitting times

For a subset of states  $A \subseteq S$ , the vector  $k^A$  of hitting times (where element  $k_i^A$  represents the expected value, starting in state  $i$  that the chain enters one of the states in the set  $A$ ) is the minimal non-negative solution to<sup>[21]</sup>

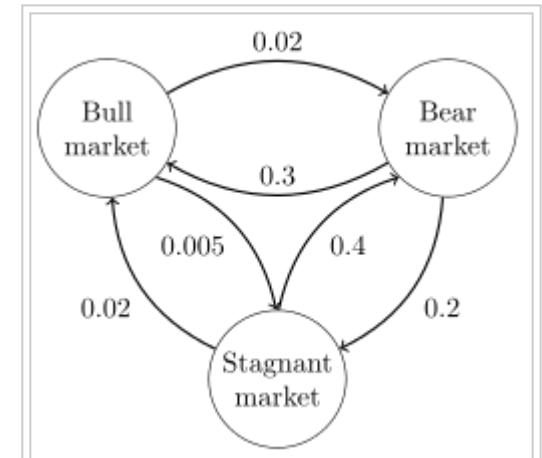
$$\begin{aligned} k_i^A &= 0 \text{ for } i \in A \\ -\sum_{j \in S} q_{ij} k_j^A &= 1 \text{ for } i \notin A. \end{aligned}$$

## Time reversal

For a CTMC  $X_t$ , the time-reversed process is defined to be  $\hat{X}_t = X_{T-t}$ . By Kelly's lemma this process has the same stationary distribution as the forward process.

A chain is said to be reversible if the reversed process is the same as the forward process. Kolmogorov's criterion states that the necessary and sufficient condition for a process to be reversible is that the product of transition rates around a closed loop must be the same in both directions.

## Embedded Markov chain



Directed graph representation of a continuous-time Markov chain describing the state of financial markets (note: numbers are made-up).



One method of finding the stationary probability distribution,  $\pi$ , of an ergodic continuous-time Markov chain,  $Q$ , is by first finding its embedded Markov chain (EMC). Strictly speaking, the EMC is a regular discrete-time Markov chain, sometimes referred to as a **jump process**. Each element of the one-step transition probability matrix of the EMC,  $S$ , is denoted by  $s_{ij}$ , and represents the conditional probability of transitioning from state  $i$  into state  $j$ . These conditional probabilities may be found by

$$s_{ij} = \begin{cases} \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases}$$

From this,  $S$  may be written as

$$S = I - (\text{diag}(Q))^{-1}Q$$

where  $I$  is the identity matrix and  $\text{diag}(Q)$  is the diagonal matrix formed by selecting the main diagonal from the matrix  $Q$  and setting all other elements to zero.

To find the stationary probability distribution vector, we must next find  $\phi$  such that

$$\phi S = \phi,$$

with  $\phi$  being a row vector, such that all elements in  $\phi$  are greater than 0 and  $\|\phi\|_1 = 1$ . From this,  $\pi$  may be found as

$$\pi = \frac{-\phi(\text{diag}(Q))^{-1}}{\|\phi(\text{diag}(Q))^{-1}\|_1}.$$

Note that  $S$  may be periodic, even if  $Q$  is not. Once  $\pi$  is found, it must be normalized to a unit vector.

Another discrete-time process that may be derived from a continuous-time Markov chain is a  **$\delta$ -skeleton**—the (discrete-time) Markov chain formed by observing  $X(t)$  at intervals of  $\delta$  units of time. The random variables  $X(0)$ ,  $X(\delta)$ ,  $X(2\delta)$ , ... give the sequence of states visited by the  $\delta$ -skeleton.

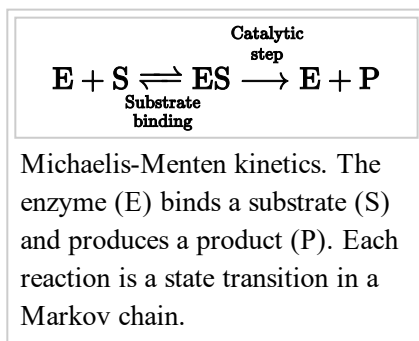
## Applications

Research has reported the application and usefulness of Markov chains in a wide range of topics such as physics, chemistry, medicine, music, game theory and sports.

## Physics

Markovian systems appear extensively in thermodynamics and statistical mechanics, whenever probabilities are used to represent unknown or unmodelled details of the system, if it can be assumed that the dynamics are time-invariant, and that no relevant history need be considered which is not already included in the state description.

## Chemistry



Markov chains and continuous-time Markov processes are useful in chemistry when physical systems closely approximate the Markov property. For example, imagine a large number  $n$  of molecules in solution in state A, each of which can undergo a chemical reaction to state B with a certain average rate. Perhaps the molecule is an enzyme, and the states refer to how it is folded. The state of any single enzyme follows a Markov chain, and since the molecules are essentially independent of each other, the number of molecules in state A or B at a time is  $n$  times the probability a given molecule is in that state.

The classical model of enzyme activity, Michaelis–Menten kinetics, can be viewed as a Markov chain, where at each time step the reaction proceeds in some direction. While Michaelis-Menten is fairly straightforward, far more complicated reaction networks can also be modeled with Markov chains.

An algorithm based on a Markov chain was also used to focus the fragment-based growth of chemicals in silico towards a desired class of compounds such as drugs or natural products.<sup>[22]</sup> As a molecule is grown, a fragment is selected from the nascent molecule as the "current" state. It is not aware of its past (i.e., it is not aware of what is already bonded to it). It then transitions to the next state when a fragment is attached to it. The transition probabilities are trained on databases of authentic classes of compounds.

Also, the growth (and composition) of copolymers may be modeled using Markov chains. Based on the reactivity ratios of the monomers that make up the growing polymer chain, the chain's composition may be calculated (e.g., whether monomers tend to add in alternating fashion or in long runs of the same monomer). Due to steric effects, second-order Markov effects may also play a role in the growth of some polymer chains.

Similarly, it has been suggested that the crystallization and growth of some epitaxial superlattice oxide materials can be accurately described by Markov chains.<sup>[23]</sup>

## Testing

Several theorists have proposed the idea of the Markov chain statistical test (MCST), a method of conjoining Markov chains to form a "Markov blanket", arranging these chains in several recursive layers ("wafering") and producing more efficient test sets—samples—as a replacement for exhaustive testing. MCSTs also have uses in temporal state-based networks; Chilukuri et al.'s paper entitled "Temporal Uncertainty Reasoning Networks for Evidence Fusion with Applications to Object Detection and Tracking" (ScienceDirect) gives a background and case study for applying MCSTs to a wider range of applications.

## Speech recognition

Hidden Markov models are the basis for most modern automatic speech recognition systems.

## Information sciences

Markov chains are used throughout information processing. Claude Shannon's famous 1948 paper *A Mathematical Theory of Communication*, which in a single step created the field of information theory, opens by introducing the concept of entropy through Markov modeling of the English language. Such idealized models can capture many of the statistical regularities of systems. Even without describing the full structure of the system perfectly, such signal models can make possible very effective data compression through entropy encoding techniques such as arithmetic coding. They also allow effective state estimation and pattern recognition. Markov chains also play an important role in reinforcement learning.

Markov chains are also the basis for hidden Markov models, which are an important tool in such diverse fields as telephone networks (which use the Viterbi algorithm for error correction), speech recognition and bioinformatics (such as in rearrangements detection<sup>[24]</sup>).

The LZMA lossless data compression algorithm combines Markov chains with Lempel-Ziv compression to achieve very high compression ratios.

## Queueing theory

Markov chains are the basis for the analytical treatment of queues (queueing theory). Agner Krarup Erlang initiated the subject in 1917.<sup>[25]</sup> This makes them critical for optimizing the performance of telecommunications networks, where messages must often compete for limited resources (such as bandwidth).<sup>[26]</sup>

Numerous queueing models use continuous-time Markov chains. For example, an M/M/1 queue is a CTMC on the non-negative integers where upward transitions from  $i$  to  $i + 1$  occur at rate  $\lambda$  according to a Poisson process and describe job arrivals, while transitions from  $i$  to  $i - 1$  (for  $i > 1$ ) occur at rate  $\mu$  (job service times are exponentially distributed) and describe completed services (departures) from the queue.

## Internet applications

The PageRank of a webpage as used by Google is defined by a Markov chain.<sup>[27]</sup> It is the probability to be at page  $i$  in the stationary distribution on the following Markov chain on all (known) webpages. If  $N$  is the number of known webpages, and a page  $i$  has  $k_i$  links to it then it has transition probability  $\frac{\alpha}{k_i} + \frac{1 - \alpha}{N}$  for all pages that are linked to and  $\frac{1 - \alpha}{N}$  for all pages that are not linked to. The parameter  $\alpha$  is taken to be about 0.85.<sup>[28]</sup>

Markov models have also been used to analyze web navigation behavior of users. A user's web link transition on a particular website can be modeled using first- or second-order Markov models and can be used to make predictions regarding future navigation and to personalize the web page for an individual user.

## Statistics

Markov chain methods have also become very important for generating sequences of random numbers to accurately reflect very complicated desired probability distributions, via a process called Markov chain Monte Carlo (MCMC). In recent years this has revolutionized the practicability of Bayesian inference methods, allowing a wide range of posterior distributions to be simulated and their parameters found numerically.

## Economics and finance

Markov chains are used in finance and economics to model a variety of different phenomena, including asset prices and market crashes. The first financial model to use a Markov chain was from Prasad *et al.* in 1974.<sup>[29]</sup> Another was the regime-switching model of James D. Hamilton (1989), in which a Markov chain is used to model switches between periods high and low GDP growth (or alternatively, economic expansions and recessions).<sup>[30]</sup> A more recent example is the Markov Switching Multifractal model of Laurent E. Calvet and Adlai J. Fisher, which builds upon the convenience of earlier regime-switching models.<sup>[31][32]</sup> It uses an arbitrarily large Markov chain to drive the level of volatility of asset returns.

Dynamic macroeconomics heavily uses Markov chains. An example is using Markov chains to exogenously model prices of equity (stock) in a general equilibrium setting.<sup>[33]</sup>

Credit rating agencies produce annual tables of the transition probabilities for bonds of different credit ratings.<sup>[34]</sup>

## Social sciences

Markov chains are generally used in describing path-dependent arguments, where current structural configurations condition future outcomes. An example is the reformulation of the idea, originally due to Karl Marx's *Das Kapital*, tying economic development to the rise of capitalism. In current research, it is common to use a Markov chain to model how once a country reaches a specific level of economic development, the configuration of structural factors, such as size of the commercial bourgeoisie, the ratio of urban to rural residence, the rate of political mobilization, etc., will generate a higher probability of transitioning from authoritarian to democratic regime.<sup>[35]</sup>

## Mathematical biology

Markov chains also have many applications in biological modelling, particularly population processes, which are useful in modelling processes that are (at least) analogous to biological populations. The Leslie matrix is one such example, though some of its entries are not probabilities (they may be greater than 1). Another example is the modeling of cell shape in dividing sheets of epithelial cells.<sup>[36]</sup> Yet another example is the state of ion channels in cell membranes.

Markov chains are also used in simulations of brain function, such as the simulation of the mammalian neocortex.<sup>[37]</sup>

## Genetics

Markov chains have been used in population genetics in order to describe the change in gene frequencies in small populations affected by genetic drift, for example in diffusion equation method described by Motoo Kimura.<sup>[38]</sup>

## Games

Markov chains can be used to model many games of chance. The children's games Snakes and Ladders and "Hi Ho! Cherry-O", for example, are represented exactly by Markov chains. At each turn, the player starts in a given state (on a given square) and from there has fixed odds of moving to certain other states (squares).

## Music

Markov chains are employed in algorithmic music composition, particularly in software such as CSound, Max and SuperCollider. In a first-order chain, the states of the system become note or pitch values, and a probability vector for each note is constructed, completing a transition probability matrix (see below). An algorithm is constructed to produce output note values based on the transition matrix weightings, which could be MIDI note values, frequency (Hz), or any other desirable metric.<sup>[39]</sup>

**1st-order matrix**

Note	A	C#	E♭
A	0.1	0.6	0.3
C#	0.25	0.05	0.7
E♭	0.7	0.3	0

**2nd-order matrix**

Notes	A	D	G
AA	0.18	0.6	0.22
AD	0.5	0.5	0
AG	0.15	0.75	0.1
DD	0	0	1
DA	0.25	0	0.75
DG	0.9	0.1	0
GG	0.4	0.4	0.2
GA	0.5	0.25	0.25
GD	1	0	0

A second-order Markov chain can be introduced by considering the current state *and* also the previous state, as indicated in the second table. Higher, *n*th-order chains tend to "group" particular notes together, while 'breaking off' into other patterns and sequences occasionally. These higher-order chains tend to generate results with a sense of phrasal structure, rather than the 'aimless wandering' produced by a first-order system.<sup>[40]</sup>

Markov chains can be used structurally, as in Xenakis's Analogique A and B.<sup>[41]</sup> Markov chains are also used in systems which use a Markov model to react interactively to music input.<sup>[42]</sup>

Usually musical systems need to enforce specific control constraints on the finite-length sequences they generate, but control constraints are not compatible with Markov models, since they induce long-range dependencies that violate the Markov hypothesis of limited memory. In order to overcome this limitation, a new approach has been proposed.<sup>[43]</sup>

## Baseball

Markov chain models have been used in advanced baseball analysis since 1960, although their use is still rare. Each half-inning of a baseball game fits the Markov chain state when the number of runners and outs are considered. During any at-bat, there are 24 possible combinations of number of outs and position of the runners. Mark Pankin shows that Markov chain models can be used to evaluate runs created for both individual players as well as a team.<sup>[44]</sup> He also discusses various kinds of strategies and play conditions: how Markov chain models have been used to analyze statistics for game situations such as bunting and base stealing and differences when playing on grass vs. astroturf.<sup>[45]</sup>

## Markov text generators

Markov processes can also be used to generate superficially real-looking text given a sample document: they are used in a variety of recreational "parody generator" software (see dissociated press, Jeff Harrison,<sup>[46]</sup> Mark V Shaney<sup>[47][48]</sup> ).

These processes are also used by spammers to inject real-looking hidden paragraphs into unsolicited email and post comments in an attempt to get these messages past spam filters.

## Bioinformatics

In the bioinformatics field, they can be used to simulate DNA sequences.<sup>[49]</sup>

## Fitting

When fitting a Markov chain to data, situations where parameters poorly describe the situation may highlight interesting trends.<sup>[50][51][52]</sup>

## History

Andrey Markov produced the first results (1906) for these processes, purely theoretically. A generalization to countably infinite state spaces was given by Kolmogorov (1936). Markov chains are related to Brownian motion and the ergodic hypothesis, two topics in physics which were important in the early years of the twentieth century. However, Markov first pursued this in 1906 as part of his argument against Pavel Nekrasov, in particular to make the case that the law of large numbers can be extended to dependent events.<sup>[53]</sup> In 1913, he applied his findings to the first 20,000 letters of Pushkin's *Eugene Onegin*.<sup>[53]</sup> By 1917, more practical application of his work was made by Erlang to obtain formulas for call loss and waiting time in telephone networks.<sup>[25]</sup>

Seneta provides an account of Markov's motivations and the theory's early development.<sup>[54]</sup> The term "chain" was used by Markov (1906) to suggest a sequence of pairwise dependent variables.<sup>[55]</sup>

## See also

- Hidden Markov model
- Markov blanket
- Markov chain geostatistics
- Markov chain mixing time
- Markov chain Monte Carlo
- Markov decision process
- Markov information source
- Markov network
- Markov process
- Quantum Markov chain
- Semi-Markov process
- Telescoping Markov chain
- Variable-order Markov model
- Brownian motion
- Dynamics of Markovian particles
- Examples of Markov chains
- Interacting particle system
- Stochastic cellular automaton
- Markov chain
- Markov decision process
- Markov model
- Random walk
- Semi-Markov process
- Markov chain approximation method

## Notes

1. Markov process (mathematics) (<http://www.britannica.com/EBchecked/topic/365797/Markov-process>) - Britannica Online Encyclopedia
2. Norris, James R. (1998). *Markov chains*. Cambridge University Press. Retrieved 2016-03-04.
3. Everitt, B.S. (2002) *The Cambridge Dictionary of Statistics*. CUP. ISBN 0-521-81099-X
4. Parzen, E. (1962) *Stochastic Processes*, Holden-Day. ISBN 0-8162-6664-6 (Table 6.1))
5. Dodge, Y. (2003) *The Oxford Dictionary of Statistical Terms*, OUP. ISBN 0-19-920613-9 (entry for "Markov chain")
6. Dodge, Y. *The Oxford Dictionary of Statistical Terms*, OUP. ISBN 0-19-920613-9
7. Meyn, S. Sean P., and Richard L. Tweedie. (2009) *Markov chains and stochastic stability*. Cambridge University Press. (Preface, p. iii)
8. Durrett, Rick (2010). *Probability: Theory and Examples* (Fourth ed.). Cambridge: Cambridge University Press. ISBN 978-0-521-76539-8.
9. S. P. Meyn and R.L. Tweedie, 2005. Markov Chains and Stochastic Stability (<http://probability.ca/MT/BOOK.pdf>)
10. Norris, J. R. (1997). "Continuous-time Markov chains I". *Markov Chains*. p. 60. doi:10.1017/CBO9780511810633.004. ISBN 9780511810633.
11. Asher Levin, David (2009). *Markov chains and mixing times*. p. 16. ISBN 978-0-8218-4739-8. Retrieved 2016-03-04.
12. Serfozo, Richard (2009), "Basics of Applied Stochastic Processes", *Probability and Its Applications*, Berlin: Springer-Verlag: 35, doi:10.1007/978-3-540-89332-5, ISBN 978-3-540-89331-8, MR 2484222
13. Florian Schmitt and Franz Rothlauf, "On the Mean of the Second Largest Eigenvalue on the Convergence Rate of Genetic Algorithms", Working Paper 1/2001, Working Papers in Information Systems, 2001. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.6191>
14. Gene H. Golub, Charles F. Van Loan, "Matrix computations", Third Edition, The Johns Hopkins University Press, Baltimore and London, 1996.
15. Franzke, Brandon; Kosko, Bart (1 October 2011). "Noise can speed convergence in Markov chains". *Physical Review E*. **84** (4). doi:10.1103/PhysRevE.84.041112.
16. Richard Durrett (19 May 2012). *Essentials of Stochastic Processes*. Springer Science & Business Media. p. 37. ISBN 978-1-4614-3615-7.

17. : A. Nielsen and M. Weber, "Computing the nearest reversible Markov chain". *Numerical Linear Algebra with Applications*, 22(3):483-499, 2015.
18. Spitzer, Frank (1970). "Interaction of Markov Processes". *Advances in Mathematics*. **5** (2): 246–290. doi:10.1016/0001-8708(70)90034-4.
19. R. L. Dobrushin; V. I. Kriukov; A. L. Toom (1978). *Stochastic Cellular Systems: Ergodicity, Memory, Morphogenesis*. ISBN 9780719022067. Retrieved 2016-03-04.
20. Doblinger, G., 1998. Smoothing of Noise AR Signals Using an Adaptive Kalman Filter. In EUSIPCO 98. pp. 781–784. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.251.3078> [Accessed January 15, 2015].
21. Norris, J. R. (1997). "Continuous-time Markov chains II". *Markov Chains*. p. 108. doi:10.1017/CBO9780511810633.005. ISBN 9780511810633.
22. Kutchukian, Peter; Lou, David; Shakhnovich, Eugene (2009). "FOG: Fragment Optimized Growth Algorithm for the de Novo Generation of Molecules occupying Druglike Chemical". *Journal of Chemical Information and Modeling*. **49** (7): 1630–1642. doi:10.1021/ci9000458. PMID 19527020.
23. Kopp, V. S.; Kaganer, V. M.; Schwarzkopf, J.; Waidick, F.; Remmele, T.; Kwasniewski, A.; Schmidbauer, M. (2011). "X-ray diffraction from nonperiodic layered structures with correlations: Analytical calculation and experiment on mixed Aurivillius films". *Acta Crystallographica Section A*. **68**: 148–155. doi:10.1107/S0108767311044874.
24. Pratas, D; Silva, R; Pinho, A; Ferreira, P (May 18, 2015). "An alignment-free method to find and visualise rearrangements between pairs of DNA sequences.". *Scientific Reports (Group Nature)*. **5** (10203): 10203. doi:10.1038/srep10203. PMID 25984837.
25. O'Connor, John J.; Robertson, Edmund F., "Markov chain", *MacTutor History of Mathematics archive*, University of St Andrews.
26. S. P. Meyn, 2007. Control Techniques for Complex Networks ([http://www.meyn.ce.ufl.edu/archive/spm\\_files/CTCN/MonographTocBib.pdf](http://www.meyn.ce.ufl.edu/archive/spm_files/CTCN/MonographTocBib.pdf)), Cambridge University Press, 2007.
27. U.S. Patent 6,285,999 (<https://www.google.com/patents/US6285999>)
28. Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999). *The PageRank Citation Ranking: Bringing Order to the Web* (Technical report). Retrieved 2016-03-04.
29. Prasad, NR; RC Ender; ST Reilly; G Nesgos (1974). "Allocation of resources on a minimized cost basis". *1974 IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes*. **13**: 402–3. doi:10.1109/CDC.1974.270470.
30. Hamilton, James (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle". *Econometrica*. *Econometrica*, Vol. 57, No. 2. **57** (2): 357–84. doi:10.2307/1912559. JSTOR 1912559.
31. Calvet, Laurent E.; Fisher, Adlai J. (2001). "Forecasting Multifractal Volatility". *Journal of Econometrics*. **105** (1): 27–58. doi:10.1016/S0304-4076(01)00069-0.
32. Calvet, Laurent; Adlai Fisher (2004). "How to Forecast long-run volatility: regime-switching and the estimation of multifractal processes". *Journal of Financial Econometrics*. **2**: 49–83. doi:10.1093/jffinec/nbh003.
33. Brennan, Michael; Xiab, Yihong. "Stock Price Volatility and the Equity Premium" (PDF). *Department of Finance, the Anderson School of Management, UCLA*.
34. A Markov Chain Example in Credit Risk Modelling Columbia University lectures ([http://www.columbia.edu/~ww2040/4106S11/MC\\_BondRating.pdf](http://www.columbia.edu/~ww2040/4106S11/MC_BondRating.pdf))
35. Acemoglu, Daron; Georgy Egorov; Konstantin Sonin (2011). "Political model of social evolution". *Proceedings of the National Academy of Sciences*. **108**: 21292–21296. doi:10.1073/pnas.1019454108.
36. Gibson, Matthew C; Patel, Ankit P.; Perrimon, Norbert; Perrimon, Norbert (2006). "The emergence of geometric order in proliferating metazoan epithelia". *Nature*. **442** (7106): 1038–1041. doi:10.1038/nature05014. PMID 16900102.
37. George, Dileep; Hawkins, Jeff (2009). Friston, Karl J., ed. "Towards a Mathematical Theory of Cortical Micro-circuits". *PLoS Comput Biol*. **5** (10): e1000532. doi:10.1371/journal.pcbi.1000532. PMC 2749218. PMID 19816557.
38. Watterson, G. (1996). "Motoo Kimura's Use of Diffusion Theory in Population Genetics". *Theoretical Population Biology* 49 (2): 154–188. doi:10.1006/tpbi.1996.0010. PMID 8813021.
39. K McAlpine; E Miranda; S Hoggar (1999). "Making Music with Algorithms: A Case-Study System". *Computer Music Journal*. **23** (2): 19–30. doi:10.1162/014892699559733.
40. Curtis Roads (ed.) (1996). *The Computer Music Tutorial*. MIT Press. ISBN 0-262-18158-4.
41. Xenakis, Iannis; Kanach, Sharon (1992) *Formalized Music: Mathematics and Thought in Composition*, Pendragon Press. ISBN 1576470792
42. Continuator (<http://www.csl.sony.fr/~pachet/>) Archived (<https://web.archive.org/web/20120713235933/http://www.csl.sony.fr/~pachet/>) July 13, 2012, at the Wayback Machine.
43. Pachet, F.; Roy, P.; Barbieri, G. (2011) "Finite-Length Markov Processes with Constraints" (<http://www.csl.sony.fr/downloads/papers/2011/pachet-11b.pdf>), *Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI*, pages 635-642, Barcelona, Spain, July 2011
44. Pankin, Mark D. "MARKOV CHAIN MODELS: THEORETICAL BACKGROUND". Retrieved 2007-11-26.
45. Pankin, Mark D. "BASEBALL AS A MARKOV CHAIN". Retrieved 2009-04-24.
46. Poet's Corner – Fieralingue ([http://www.fieralingue.it/modules.php?name=Content&pa=list\\_pages\\_categories&cid=111](http://www.fieralingue.it/modules.php?name=Content&pa=list_pages_categories&cid=111)) Archived ([https://web.archive.org/web/20101206043430/http://www.fieralingue.it/modules.php?name=Content&pa=list\\_pages\\_categories&cid=111](https://web.archive.org/web/20101206043430/http://www.fieralingue.it/modules.php?name=Content&pa=list_pages_categories&cid=111)) December 6, 2010, at the Wayback Machine.



47. Kenner, Hugh; O'Rourke, Joseph (November 1984). "A Travesty Generator for Micros". *BYTE*. **9** (12): 129–131, 449–469.
48. Hartman, Charles (1996). *Virtual Muse: Experiments in Computer Poetry*. Hanover, NH: Wesleyan University Press. ISBN 0-8195-2239-2.
49. Pratas, Diogo; Bastos, Carlos; Pinho, Armando; Neves, Antonio; Matos, Luis (June 2011). *DNA synthetic sequences generation using multiple competing Markov models. Statistical Signal Processing Workshop (SSP), 2011 IEEE*. **9** (12). pp. 133–136. doi:10.1109/SSP.2011.5967639.
50. Avery, P. J.; Henderson, D. A. (1999). "Fitting Markov Chain Models to Discrete State Series Such as DNA Sequences". *Journal of the Royal Statistical Society*. **48** (1): 53–61. doi:10.1111/1467-9876.00139. JSTOR 2680818.
51. Shmilovici A. & Ben-Gal I. (2007). "Using a VOM Model for Reconstructing Potential Coding Regions in EST Sequences," (PDF). *Journal of Computational Statistics*, vol. 22, no. 1, 49–69. Retrieved 2016-03-04.
52. [http://www.eng.tau.ac.il/~bengal/VOM\\_EST.pdf](http://www.eng.tau.ac.il/~bengal/VOM_EST.pdf)
53. Hayes, Brian (March–April 2013). "First Links in the Markov Chain". *American Scientist*. **101**: 92–97.
54. Seneta, E. (1996). "Markov and the Birth of Chain Dependence Theory". *International Statistical Review*. **64** (3): 255–263. doi:10.2307/1403785. JSTOR 1403785.
55. Upton, G.; Cook, I. (2008). *Oxford Dictionary of Statistics*. OUP. ISBN 978-0-19-954145-4.

## References

- A.A. Markov. "Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga". *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2-ya seriya, tom 15, pp. 135–156, 1906.
- A.A. Markov. "Extension of the limit theorems of probability theory to a sum of variables connected in a chain". reprinted in Appendix B of: R. Howard. *Dynamic Probabilistic Systems, volume 1: Markov Chains*. John Wiley and Sons, 1971.
- Classical Text in Translation: A. A. Markov, An Example of Statistical Investigation of the Text Eugene Onegin Concerning the Connection of Samples in Chains, trans. David Link. *Science in Context* 19.4 (2006): 591–600. Online: <http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=637500>
- Leo Breiman. *Probability*. Original edition published by Addison-Wesley, 1968; reprinted by Society for Industrial and Applied Mathematics, 1992. ISBN 0-89871-296-3. (See Chapter 7.)
- J.L. Doob. *Stochastic Processes*. New York: John Wiley and Sons, 1953. ISBN 0-471-52369-0.
- S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. London: Springer-Verlag, 1993. ISBN 0-387-19832-6. online: [https://netfiles.uiuc.edu/meyn/www/spm\\_files/book.html](https://netfiles.uiuc.edu/meyn/www/spm_files/book.html) . Second edition to appear, Cambridge University Press, 2009.
- S. P. Meyn. *Control Techniques for Complex Networks*. Cambridge University Press, 2007. ISBN 978-0-521-88441-9. Appendix contains abridged Meyn & Tweedie. online: [https://netfiles.uiuc.edu/meyn/www/spm\\_files/CTCN/CTCN.html](https://netfiles.uiuc.edu/meyn/www/spm_files/CTCN/CTCN.html)Booth, Taylor L. (1967). *Sequential Machines and Automata Theory* (1st ed.). New York: John Wiley and Sons, Inc. Library of Congress Card Catalog Number 67-25924. Extensive, wide-ranging book meant for specialists, written for both theoretical computer scientists as well as electrical engineers. With detailed explanations of state minimization techniques, FSMs, Turing machines, Markov processes, and undecidability. Excellent treatment of Markov processes pp. 449ff. Discusses Z-transforms, D transforms in their context.
- Kemeny, John G.; Hazleton Mirkil; J. Laurie Snell; Gerald L. Thompson (1959). *Finite Mathematical Structures* (1st ed.). Englewood Cliffs, N.J.: Prentice-Hall, Inc. Library of Congress Card Catalog Number 59-12841. Classical text. cf Chapter 6 *Finite Markov Chains* pp. 384ff.
- E. Nummelin. "General irreducible Markov chains and non-negative operators". Cambridge University Press, 1984, 2004. ISBN 0-521-60494-X
- Seneta, E. *Non-negative matrices and Markov chains*. 2nd rev. ed., 1981, XVI, 288 p., Softcover Springer Series in Statistics. (Originally published by Allen & Unwin Ltd., London, 1973) ISBN 978-0-387-29765-1
- Kishor S. Trivedi, *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*, John Wiley & Sons, Inc. New York, 2002. ISBN 0-471-33341-7.
- K.S.Trivedi and R.A.Sahner, *SHARPE at the age of twenty-two*, vol. 36, no. 4, pp.-52-57, ACM SIGMETRICS Performance Evaluation Review, 2009.
- R.A.Sahner, K.S.Trivedi and A. Puliafito, *Performance and reliability analysis of computer systems: an example-based approach using the SHARPE software package*, Kluwer Academic Publishers, 1996. ISBN 0-7923-9650-2.
- G.Bolch, S.Greiner, H.de Meer and K.S.Trivedi, *Queueing Networks and Markov Chains*, John Wiley, 2nd edition, 2006. ISBN 978-0-7923-9650-5.

## External links

- Introduction to Markov Chains ([https://www.youtube.com/watch?v=o-jdJxXL\\_W4](https://www.youtube.com/watch?v=o-jdJxXL_W4)) on YouTube
- Hazewinkel, Michiel, ed. (2001), "Markov chain", *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Techniques to Understand Computer Simulations: Markov Chain Analysis (<http://jasss.soc.surrey.ac.uk/12/1/6.html>)
- Markov Chains chapter in American Mathematical Society's introductory probability book ([http://www.dartmouth.edu/~chance/teaching\\_aids/books\\_articles/probability\\_book/Chapter11.pdf](http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/Chapter11.pdf))(pdf)
- A beautiful visual explanation of Markov Chains (<http://setosa.io/blog/2014/07/26/markov-chains/index.html>)
- Chapter 5: Markov Chain Models (<http://www.math.rutgers.edu/courses/338/coursenotes/chapter5.pdf>)
- Making Sense and Nonsense of Markov Chains (<http://rarlindseymash.com/posts/2009-11-21-making-sense-and-nonsense-of-markov-chains>)

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Markov\\_chain&oldid=740916871](https://en.wikipedia.org/w/index.php?title=Markov_chain&oldid=740916871)"

Categories: Markov processes | Markov models | Graph theory

- 
- This page was last modified on 24 September 2016, at 04:38.
  - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.