

54

Markov Chains

Beatrice Meini
Università di Pisa, Italy

54.1 Basic Concepts	54-1
54.2 Irreducible Classes	54-5
54.3 Classification of the States	54-7
54.4 Finite Markov Chains	54-10
54.5 Censoring	54-11
54.6 Numerical Methods	54-12
References	54-15

Markov chains are encountered in several applications arising in different contexts, and model many real problems which evolve in time. Throughout, we denote by $P[X = j]$ the probability that the random variable X takes the value j , and by $P[X = j|Y = i]$ the conditional probability that X takes the value j , given that the random variable Y takes the value i . Moreover, we denote by $E[X]$ the expected value of the random variable X and by $E[X|A]$ the conditional expectation of X , given the event A .

54.1 Basic Concepts

Definitions:

Given a denumerable set E , a **discrete stochastic process** on E is a family $\{X_t : t \in T\}$ of random variables X_t indexed by some denumerable set T and with values in E , i.e., $X_t \in E$ for all $t \in T$. Here, E is the **state space**, and T is the **time space**.

The discrete stochastic process $\{X_n : n \in \mathbb{N}\}$ is a **Markov chain** if

$$P[X_{n+1} = j_{n+1} | X_0 = j_0, X_1 = j_1, \dots, X_n = j_n] = P[X_{n+1} = j_{n+1} | X_n = j_n], \quad (54.1)$$

for any $(n+2)$ -tuple of states $\{j_0, \dots, j_{n+1}\} \in E$, and for all time $n \in \mathbb{N}$.

A Markov chain $\{X_n : n \in \mathbb{N}\}$ is **homogeneous** if

$$P[X_{n+1} = j | X_n = i] = P[X_1 = j | X_0 = i], \quad (54.2)$$

for all states $i, j \in E$ and for all time $n \in \mathbb{N}$.

Given a homogeneous Markov chain $\{X_n : n \in \mathbb{N}\}$ we define the **transition matrix** of the Markov chain to be the matrix $P = [p_{i,j}]_{i,j \in E}$ such that

$$p_{i,j} = P[X_1 = j | X_0 = i], \quad \text{for all } i, j \text{ in } E.$$

Throughout, unless differently specified, for the sake of notational simplicity we will indicate with the term **Markov chain** a homogeneous Markov chain.

A **finite Markov chain** is a Markov chain with finite space state E ; an **infinite Markov chain** is a Markov chain with infinite space state E .

A row vector $\pi = (\pi_i)_{i \in E}$ such that

$$\pi P = \pi \quad (54.3)$$

is an **invariant vector**.

If the invariant vector π is such that

$$\pi_i \geq 0 \text{ for all } i, \quad \text{and} \quad \sum_{i \in E} \pi_i = 1, \quad (54.4)$$

then π is an **invariant probability vector**, or **stationary distribution**. The definition of stationary distribution extends the definition given in Section 9.4 to the case of an infinite matrix P .

Facts:

1. The Markov property in Equation (54.1) means that the state X_n of the system at time n is sufficient to determine which state might be occupied at time $n + 1$, and the past history X_0, X_1, \dots, X_{n-1} does not influence the future state at time $n + 1$.
2. In a homogeneous Markov chain, the property in Equation (54.2) means that the laws which govern the evolution of the system are independent of the time n ; therefore, the evolution of the Markov chain is ruled by the transition matrix P , whose (i, j) th entry represents the probability to change from state i to state j in one time unit.
3. The number of rows and columns of the transition matrix P is equal to the cardinality of E . In particular, if the set E is finite, P is a finite matrix (see Examples 1 and 5); if the set E is infinite, P is an infinite matrix, i.e., a matrix with an infinite number of rows and columns (see Examples 2–4).
4. The matrix P is row stochastic, i.e., it is a matrix with nonnegative entries such that $\sum_{j \in E} p_{i,j} = 1$ for any $i \in E$, i.e., the sum of the entries on each row is equal to 1 (see Section 9.4).
5. If $|E| < \infty$, the matrix P has spectral radius 1. Moreover, the vector $\mathbf{1}_{|E|}$ is a right eigenvector of P corresponding to the eigenvalue 1; any nonzero invariant vector is a left eigenvector of P corresponding to the eigenvalue 1; the invariant probability vector π is a nonnegative left eigenvector of P corresponding to the eigenvalue 1, normalized so that $\pi \mathbf{1}_{|E|} = 1$ (Fact 1 of Section 9.4).
6. In the analysis of Markov chains, we may encounter matrix products $A = BC$, where $B = [b_{i,j}]_{i,j \in E}$ and $C = [c_{i,j}]_{i,j \in E}$ are nonnegative matrices such that $\sum_{j \in E} b_{i,j} \leq 1$ and $\sum_{j \in E} c_{i,j} \leq 1$ for any $i \in E$ (see, for instance, Fact 7 below). The (i, j) th entry of A , given by $a_{i,j} = \sum_{h \in E} b_{i,h} c_{h,j}$, is well defined also if the set E is infinite, since $0 \leq \sum_{h \in E} b_{i,h} c_{h,j} \leq \sum_{h \in E} b_{i,h} \leq 1$. Moreover, A is a nonnegative matrix such that $\sum_{j \in E} a_{i,j} \leq 1$ for any $i \in E$. Indeed, $\sum_{j \in E} a_{i,j} = \sum_{j \in E} \sum_{h \in E} b_{i,h} c_{h,j} = \sum_{h \in E} b_{i,h} \sum_{j \in E} c_{h,j} \leq \sum_{h \in E} b_{i,h} \leq 1$. Similarly, the product $\mathbf{v} = \mathbf{u}B$, where $\mathbf{u} = (u_i)_{i \in E}$ is a nonnegative row vector such that $\sum_{i \in E} u_i \leq 1$, is well defined also in the case where E is infinite; moreover, $\mathbf{v} = (v_i)_{i \in E}$ is a nonnegative vector such that $\sum_{i \in E} v_i \leq 1$.
7. [Nor99, Theorem 1.1.3] The dynamic behavior of a homogeneous Markov chain is completely characterized by the transition matrix P :

$$P[X_{n+k} = j | X_n = i] = (P^k)_{i,j},$$

for all times $n \geq 0$, all intervals of time $k \geq 0$, and all pairs of states i and j in E .

8. In addition to the system dynamics, one must choose the starting point X_0 . Let $\pi^{(0)} = (\pi_i^{(0)})_{i \in E}$ be a probability distribution on E , i.e., a nonnegative row vector such that the sum of its components is equal to one. Assume that $\pi_i^{(0)} = P[X_0 = i]$, and define the row vector $\pi^{(n)} = (\pi_i^{(n)})_{i \in E}$ to be the probability vector of the Markov chain at time $n \geq 1$, that is, $\pi_i^{(n)} = P[X_n = i | X_0]$. Then

$$\pi^{(n+1)} = \pi^{(n)} P, \quad n \geq 0, \quad (54.5)$$

$$\pi^{(n)} = \pi^{(0)} P^n, \quad n \geq 0. \quad (54.6)$$

9. If the initial distribution $\pi^{(0)}$ coincides with the invariant probability vector π , then $\pi^{(n)} = \pi$ for any $n \geq 0$.
10. In certain applications (see for instance, Example 5) we are interested in the asymptotic behavior of the Markov chain. In particular, we would like to compute, if it exists, the vector $\lim_{n \rightarrow \infty} \pi^{(n)}$. From Equation (54.5) of Fact 8 one deduces that, if such limit exists, it coincides with the invariant probability vector, i.e., $\lim_{n \rightarrow \infty} \pi^{(n)} = \pi$.

Examples:

1. *Random walk on $\{0, 1, \dots, k\}$:* Consider a particle which moves on the interval $[0, k]$ in unit steps at integer instants of time; let $X_n \in \{0, 1, \dots, k\}$, $n \geq 0$, be the position of the particle at time n and let $0 < p < 1$. Assume that, if the particle is in the open interval $(0, k)$, at the next unit time it will move to the right with probability p and to the left with probability $q = 1 - p$; if the particle is in position 0, it will move to the right with probability 1; if the particle is in position k , it will move to the left with probability 1. Clearly, the discrete stochastic process $\{X_n : n \in \mathbb{N}\}$ is a homogeneous Markov chain with space state the set $E = \{0, 1, \dots, k\}$. The transition matrix is the $(k+1) \times (k+1)$ matrix $P = [p_{i,j}]_{i,j=0,\dots,k}$, given by

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

From Equation (54.6) of Fact 8 one has that, if the particle is in position 0 at time 0, the probability vector at time $n = 10$ is the vector

$$[\pi_0^{(10)}, \pi_1^{(10)}, \dots, \pi_k^{(10)}] = [1, 0, \dots, 0] P^{10}.$$

2. *Random walk on \mathbb{N} :* If we allow the particle to move on \mathbb{N} , we have a homogeneous Markov

chain with state space the set $E = \mathbb{N}$. The transition matrix is semi-infinite and is given by

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

3. *Random walk on \mathbb{Z}* : If we allow the particle to move on \mathbb{Z} , we still have a homogeneous Markov chain, with state space the set $E = \mathbb{Z}$. The transition matrix is bi-infinite and is given by

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

4. *A simple queueing system* [BLM05, Example 1.3]: Simple queues consist of one server which attends to one customer at a time, in order of their arrivals. We assume that time is discretized into intervals of length one, that a random number of customers join the system during each interval, that customers do not leave the queue, and that the server removes one customer from the queue at the end of each interval, if there is any. Defining α_n as the number of new arrivals during the interval $[n-1, n)$ and X_n as the number of customers in the system at time n , we have

$$X_{n+1} = \begin{cases} X_n + \alpha_{n+1} - 1 & \text{if } X_n + \alpha_{n+1} \geq 1 \\ 0 & \text{if } X_n + \alpha_{n+1} = 0. \end{cases}$$

If $\{\alpha_n\}$ is a collection of independent random variables, then X_{n+1} is conditionally independent of X_0, \dots, X_{n-1} if X_n is known. If, in addition, the α_n 's are identically distributed, then $\{X_n\}$ is homogeneous. The state space is \mathbb{N} and the transition matrix is

$$P = \begin{bmatrix} q_0 + q_1 & q_2 & q_3 & q_4 & \dots \\ q_0 & q_1 & q_2 & q_3 & \ddots \\ & q_0 & q_1 & q_2 & \ddots \\ 0 & & \ddots & \ddots & \ddots \end{bmatrix},$$

where q_i is the probability $P[\alpha = i]$ that i new customers join the queue during a unit time interval, α denoting any of the identically distributed random variables α_n . Markov chains

having transition matrix of the form

$$P = \begin{bmatrix} B_1 & B_2 & B_3 & B_4 & \dots \\ A_0 & A_1 & A_2 & A_3 & \ddots \\ & A_0 & A_1 & A_2 & \ddots \\ 0 & & \ddots & \ddots & \ddots \end{bmatrix},$$

where A_i , B_{i+1} , $i \geq 0$, are nonnegative $k \times k$ matrices, are called M/G/1-type Markov chains, and model a large variety of queuing problems. (See [Neu89], [LR99], and [BLM05].)

5. *Search engines* (see Section 63.5, [PBM99], [ANT02] and [Mol04, Section 2.11]): PageRank is used by Google to sort, in order of relevance, the pages on the Web that match the query of the user. From the Web page *Google Technology* at <http://www.google.com/technology/>: “The heart of our software is PageRankTM, a system for ranking Web pages developed by our founders Larry Page and Sergey Brin at Stanford University. And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our Web search tools.” Surfing on the Web is seen as a random walk, where either one starts from a Web page and goes from one page to the next page by randomly following a link (if any), or one simply chooses a random page from the Web. Let E be the set of Web pages that can be reached by following a sequence of hyperlinks starting from a given Web page. If k is the number of Web pages, i.e., $k = |E|$, the connectivity matrix is defined as the $k \times k$ matrix $G = [g_{i,j}]$ such that $g_{i,j} = 1$ if there is a hyperlink from page i to page j , and zero otherwise. Let $r_i = \sum_j g_{i,j}$ and $c_i = \sum_j g_{j,i}$ be the row and column sums of G ; the quantities r_i and c_i are called the out-degree and the in-degree, respectively, of the page i . Let $0 < q < 1$ and let $P = [p_{i,j}]$ be the $k \times k$ stochastic matrix such that $p_{i,j} = qg_{i,j}/r_i + (1 - q)/k$, $i, j = 1, \dots, k$. The value q is the probability that the random walk on E follows a link and, therefore, $1 - q$ is the probability that an arbitrary page is chosen. The matrix P is the transition matrix of the Markov chain that models the random walk on E . The importance of a Web page is related to the probability to reach such page during this random walk as the time tends to infinity. Therefore, Google’s PageRank is determined by the invariant probability vector π of P : The larger the value of an entry π_i of π , the higher the relevance of the i th Web page in the set E .

54.2 Irreducible Classes

Some of the Definitions and Facts given in this section extend the corresponding Definitions and Facts of Chapter 9 and Chapter 29. Indeed, in these chapters, it is assumed that the matrices have a finite size, while in the framework of Markov chains we may encounter infinite matrices.

Definitions:

The **transition graph** of a Markov chain with transition matrix P is the digraph of P , $\Gamma(P)$. That is, the digraph defined as follows: To each state in E there corresponds a vertex of the digraph and one defines a directed arc from vertex i to vertex j for each pair of states such that $p_{i,j} > 0$. More information about digraphs can be found in Chapter 9 and Chapter 29.

A **closed walk** in a digraph is a walk in which the first vertex equals the last vertex.

State i **leads** to j (or i **has access** to j) if there is a walk from i to j in the transition graph.

States i and j **communicate** if i leads to j and j leads to i .

A Markov chain is called **irreducible** if the transition graph is strongly connected, i.e., if all the states communicate. A Markov chain is called **reducible** if it is not irreducible.

The strongly connected components of the transition graph are the **communicating classes of states**, or of the Markov chain. Communicating classes are also called **access equivalence classes** or **irreducible classes**.

A communicating class C is a **final class** if for every state i in C , there is no state j outside of C such that i leads to j . If, on the contrary, there is a state in C that leads to some state outside of C , the class is a **passage class**. A single state that forms a final class by itself is **absorbing**. In Section 9.4, passage classes are called transient classes, and final classes are called ergodic classes; in fact, transient and ergodic states of Markov chains will be introduced in Section 54.3, and for finite Markov chains the states in a passage class are transient, and the states in a final class are ergodic, cf. Section 54.4.

A state i is **periodic** with period $\delta \geq 2$ if all closed walks through i in the transition graph have a length that is a multiple of δ . A state i is **aperiodic** if it is not periodic.

A Markov chain is **periodic** with period δ if all states are periodic and have the same period δ .

Facts:

1. A Markov chain is irreducible if and only if the transition matrix P is irreducible, i.e., if P is not permutation similar to a block triangular matrix, cf. Section 9.2, Section 27.1.
2. If we adopt the convention that each state communicates with itself, then the relation *communicates* is an equivalence relation and the communicating classes are the equivalence classes of this relation, cf. Section 9.1.
3. A Markov chain is irreducible if and only if the states form one single communicating class, cf. Section 9.1, Section 9.2.
4. If a Markov chain with transition matrix P has $K \geq 2$ communicating classes, denoted by C_1, C_2, \dots, C_K , then the states may be permuted so that the transition matrix $P' = \Pi P \Pi^T$ associated with the permuted states is block triangular:

$$P' = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,K} \\ & P_{2,2} & \ddots & \vdots \\ & & \ddots & P_{K-1,K} \\ 0 & & & P_{K,K} \end{bmatrix}$$

where $P_{i,j}$ is the submatrix of transition probabilities from the states of C_i to C_j , the diagonal blocks are irreducible square matrices, and Π is the permutation matrix associated with the rearrangement, cf. Section 9.2.

5. [Çin75, Theorem (3.16), Chap. 5] Periodicity is a class property and all states in a communicating class have the same period. Thus, for irreducible Markov chains, either all states are aperiodic, or all have the same period δ , which we may call the period of the Markov chain itself.

Examples:

1. Figure 1 is the transition graph associated with the Markov chain on $E = \{1, 2, 3\}$ with

transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}.$$

The two sets $C_1 = \{1\}$ and $C_2 = \{2, 3\}$ are two communicating classes. C_2 is a passage class, while C_1 is a final class, therefore state 1 is absorbing.

2. Figure 2 is the transition graph associated with the Markov chain on $E = \{k \in \mathbb{N} : k \geq 1\}$ with transition matrix having following structure:

$$P = \begin{bmatrix} 0 & * & & & & 0 \\ 0 & 0 & * & & & \\ * & 0 & 0 & * & & \\ 0 & 0 & 0 & 0 & * & \\ 0 & 0 & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots \\ 0 & 0 & 0 & 0 & * & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix},$$

where “*” denotes a nonzero element. This is an example of a periodic irreducible Markov chain of period 3.

54.3 Classification of the States

Definitions:

Let T_j be the time of the first visit to j , without taking into account the state at time 0, i.e.,

$$T_j = \min\{n \geq 1 : X_n = j\}.$$

Define

$$f_j = P[T_j < \infty | X_0 = j],$$

that is, f_j is the probability that, starting from j , the Markov chain returns to j in a finite time.

A state $j \in E$ is **transient** if $f_j < 1$.

A state $j \in E$ is **recurrent** if $f_j = 1$. A recurrent state $j \in E$ is **positive recurrent** if the expected return time $E[T_j | X_0 = j]$ is finite; it is **null recurrent** if the expected return time $E[T_j | X_0 = j]$ is infinite. Positive recurrent states are also called **ergodic**.

A Markov chain is **positive/null recurrent** or **transient** if all its states are positive/null recurrent or transient, respectively.

A **regular Markov chain** is a positive recurrent Markov chain that is aperiodic.

The matrix $R = \sum_{n=0}^{\infty} P^n$ is the **potential matrix** of the Markov chain.

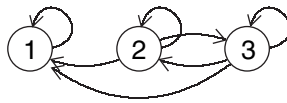


FIGURE 54.1 Transition graph of the Markov chain of Example 1.



FIGURE 54.2 Graph of an infinite irreducible periodic Markov chain of period 3.

Facts:

1. Transient states may be visited only a finite number of times by the Markov chain. On the contrary, once the Markov chain has visited one recurrent state, it will return to it over and over again; if j is null recurrent the expected time between two successive visits to j is infinite.
2. [Çin75, Cor. (2.13), Chap. 5] For the potential matrix $R = [r_{i,j}]$ one has $r_{j,j} = 1/(1 - f_j)$, where we set $1/0 = \infty$.
3. [Çin75, Sec. 3, Chap. 5] The (j,j) th entry of R is the expected number of returns of the Markov chain to state j . A state j is recurrent if and only if $r_{j,j} = \infty$. A state j is transient if and only if $r_{j,j} < \infty$.
4. [Nor99, Theorems 1.5.4, 1.5.5, 1.7.7] The nature of a state is a class property. More specifically, the states in a passage class are transient; in a final class the states are either all positive recurrent, or all null recurrent, or all transient.
5. [Nor99, Theorem 1.5.6] If a final class contains a finite number of states only, then all its states are positive recurrent.
6. From Facts 4 and 5 one has that, for a communicating class with a finite number of states, either all the states are transient or all the states are positive recurrent.
7. From Fact 4 above, if a Markov chain is irreducible, the states are either all positive recurrent, or all null recurrent, or all transient. Therefore, an irreducible Markov chain is either positive recurrent or null recurrent or transient.
8. [Nor99, Secs. 1.7 and 1.8] Assume that the Markov chain is irreducible. The Markov chain is positive recurrent if and only if there exists a strictly positive invariant probability vector, that is, a row vector $\pi = (\pi_i)_{i \in E}$ such that $\pi_i > 0$ that satisfies Equations (54.3), (54.4) of Section 54.1. The invariant vector π is unique among nonnegative vectors, up to a multiplicative constant.
9. [Nor99, Secs. 1.7 and 1.8], [BLM05, Sec. 1.5] If the Markov chain is irreducible and null recurrent, there exists a strictly positive invariant vector, unique up to a multiplicative constant, such that the sum of its elements is not finite. Thus, there always exists an invariant vector for the transition matrix of a recurrent Markov chain. Some transient Markov chains also have an invariant vector (with infinite sum of the entries, like in the null recurrent case) but some do not.
10. [Çin75, Theorem (3.2), Chap. 5], [Nor99, Secs. 1.7 and 1.8] If j is a transient or null recurrent state, then for any $i \in E$, $\lim_{n \rightarrow \infty} (P^n)_{i,j} = 0$. If j is a positive recurrent and aperiodic state, then $\lim_{n \rightarrow \infty} (P^n)_{j,j} > 0$. If j is periodic with period δ , then $\lim_{n \rightarrow \infty} (P^{n\delta})_{j,j} > 0$.
11. [Nor99, Secs. 1.7 and 1.8] Assume that the Markov chain is irreducible, aperiodic, and positive recurrent. Then $\lim_{n \rightarrow \infty} (P^n)_{i,j} = \pi_j > 0$ for all j , independently of i , where $\pi = (\pi_i)_{i \in E}$ is the stationary distribution.
12. Facts 3, 4, 6, and 10 provide a criterium to classify the states. First identify the communicating classes. If a communicating class contains a finite number of states, the states are all positive recurrent if the communicating class is final; the states are all transient if the communicating class is a passage class. If a communicating class has infinitely many states, we apply Fact 3 to determine if they are recurrent or transient; for recurrent states we use Fact 10 to determine if they are null or positive recurrent.

Examples:

1. Let P be the transition matrix of Example 1 of Section 54.2. Observe that with positive probability the Markov chain moves from state 2 or 3 to state 1, and when the Markov chain will be in state 1, it will remain there forever. Indeed, according to Facts 4 and 5, states 2 and 3 are transient since they belong to a passage class, and state 1 is positive recurrent since it is absorbing. Moreover, if we partition the matrix P into the 2×2 block matrix

$$P = \begin{bmatrix} 1 & 0 \\ A & B \end{bmatrix},$$

where

$$A = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{4} \end{bmatrix}, \quad B = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} \end{bmatrix},$$

we may easily observe that

$$P^n = \begin{bmatrix} 1 & 0 \\ \sum_{i=0}^{n-1} B^i A & B^n \end{bmatrix}.$$

Since $\|B\|_1 = \frac{5}{6}$, then $\rho(B) < 1$, and therefore $\lim_{n \rightarrow \infty} B^n = 0$ and $\sum_{n=0}^{\infty} B^n = (I - B)^{-1}$. A simple computation leads to

$$\lim_{n \rightarrow \infty} P^n = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad R = \sum_{n=0}^{\infty} P^n = \begin{bmatrix} \infty & 0 & 0 \\ \infty & \frac{9}{4} & 1 \\ \infty & \frac{3}{2} & 2 \end{bmatrix},$$

in accordance with Facts 3 and 10.

2. [BLM05, Ex. 1.19] The transition matrix

$$P = \begin{bmatrix} 0 & 1 & & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & \\ & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & & \ddots & \ddots & \ddots \end{bmatrix}$$

is irreducible, and $\pi = [\frac{1}{2}, 1, 1, \dots]$ is an invariant vector. The vector π has “infinite” mass, that is, the sum of its components is infinite. In fact, the Markov chain is actually null recurrent (see [BLM05, Sec. 1.5]).

3. [BLM05, Ex. 1.20] For the transition matrix

$$P = \begin{bmatrix} 0 & 1 & & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & \\ & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & & \ddots & \ddots & \ddots \end{bmatrix}$$

one has $\pi P = \pi$ with $\pi = [1, 4, 12, 36, 108, \dots]$. The vector π has unbounded elements. In this case the Markov chain is transient.

4. [BLM05, Ex. 1.21] For the transition matrix

$$P = \begin{bmatrix} 0 & 1 & & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & \\ & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & & \ddots & \ddots & \ddots \end{bmatrix}$$

one has $\pi P = \pi$ with $\pi = \frac{4}{3}[\frac{1}{4}, \frac{1}{3}, \frac{1}{9}, \frac{1}{27}, \dots]$ and $\sum_i \pi_i = 1$. In this case the Markov chain is positive recurrent by Fact 8.

54.4 Finite Markov Chains

The transition matrix of a finite Markov chain is a finite dimensional stochastic matrix; the reader is advised to consult Section 9.4 for properties of finite stochastic matrices.

Definitions:

A $k \times k$ matrix A is **weakly cyclic of index δ** if there exists a permutation matrix Π such that $A' = \Pi A \Pi^T$ has the block form

$$A' = \begin{bmatrix} 0 & 0 & \dots & 0 & A_{1,\delta} \\ A_{2,1} & 0 & \ddots & & 0 \\ 0 & A_{3,2} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \dots & 0 & A_{\delta,\delta-1} & 0 \end{bmatrix},$$

where the zero diagonal blocks are square.

Facts:

1. If P is the transition matrix of a finite Markov chain, there exists an invariant probability vector π . If P is irreducible, the vector π is strictly positive and unique, cf. Section 9.4.
2. For a finite Markov chain no state is null recurrent, and not all states are transient. The states belonging to a final class are positive recurrent, and the states belonging to a passage class are transient, cf. Section 9.4.
3. [BP94, Theorem (3.9), Chap. 8] Let P be the transition matrix of a finite Markov chain. The Markov chain is
 - (a) Positive recurrent if and only if P is irreducible.
 - (b) Regular if and only if P is primitive.
 - (c) Periodic if and only if P is irreducible and periodic.
4. [BP94, Theorem (3.16), Chap. 8] Let P be the transition matrix of a finite irreducible Markov chain. Then the Markov chain is periodic if and only if P is a weakly cyclic matrix.
5. If P is the transition matrix of a regular Markov chain, then there exists $\lim_{n \rightarrow \infty} P^n = \mathbf{1}\pi$, where π is the probability invariant vector. If the Markov chain is periodic, the sequence $\{P^n\}_{n \geq 0}$ is bounded, but not convergent.
6. [Mey89] Assume that $E = \{1, 2, \dots, k\}$ and that P is irreducible. Let $1 \leq m \leq k-1$ and $\alpha = \{1, \dots, m\}$, $\beta = \{m+1, \dots, k\}$. Then
 - (a) The matrix $I - P[\alpha]$ is an M -matrix.
 - (b) The matrix $P' = P[\alpha] + P[\alpha, \beta](I - P[\beta])^{-1}P[\beta, \alpha]$, such that $I - P'$ is the Schur complement of $I - P[\beta]$ in the matrix

$$\begin{bmatrix} I - P[\alpha] & -P[\alpha, \beta] \\ -P[\beta, \alpha] & I - P[\beta] \end{bmatrix},$$

is a stochastic irreducible matrix.

- (c) If we partition the invariant probability vector π as $\pi = [\pi_\alpha, \pi_\beta]$, with $\pi_\alpha = (\pi_i)_{i \in \alpha}$ and $\pi_\beta = (\pi_i)_{i \in \beta}$, we have $\pi_\alpha P' = \pi_\alpha$, $\pi_\beta = \pi_\alpha P[\alpha, \beta](I - P[\beta])^{-1}$.

Examples:

1. The matrix

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

is the transition matrix of a periodic Markov chain of period 2. In fact, P is an irreducible matrix of period 2.

2. The Markov chain with transition matrix

$$P = \begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ 1 & 0 \end{bmatrix}$$

is regular. In fact, $P^2 > 0$, i.e., P is primitive.

3. Let

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

be the transition matrix of a Markov chain. The matrix P is irreducible and aperiodic, therefore the Markov chain is regular. The vector $\pi = \frac{1}{13}[4, 4, 3, 2]$ is the stationary distribution. According to Fact 5, one has

$$\lim_{n \rightarrow \infty} P^n = \frac{1}{13} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 4 & 4 & 3 & 2 \end{bmatrix}.$$

54.5 Censoring

Definitions:

Partition the state space E into two disjoint subsets, α and β , and denote by $\{t_0, t_1, t_2, \dots\}$ the time when the Markov chain visits the set α :

$$t_0 = \min\{n \geq 0 : X_n \in \alpha\}, \quad t_{k+1} = \min\{n \geq t_k + 1 : X_n \in \alpha\},$$

for $k \geq 0$. The **censored process** restricted to the subset α is the sequence $\{Y_n\}_{n \geq 0}$, where $Y_n = X_{t_n}$, of successive states visited by the Markov chain in α .

Facts:

1. [BLM05, Sec. 1.6] The censored process $\{Y_n\}$ is a Markov chain.

2. [BLM05, Sec. 1.6] Arrange the states so that the transition matrix can be partitioned as

$$P = \begin{bmatrix} P[\alpha] & P[\alpha, \beta] \\ P[\beta, \alpha] & P[\beta] \end{bmatrix}.$$

Then the transition matrix of $\{Y_n\}$ is

$$P' = P[\alpha] + \sum_{n=0}^{+\infty} (P[\alpha, \beta]P[\beta]^n P[\beta, \alpha])$$

provided that $\sum_{n=0}^{+\infty} (P[\alpha, \beta]P[\beta]^n P[\beta, \alpha])$ is convergent. If the series $S' = \sum_{n=0}^{+\infty} P[\beta]^n$ is convergent, then we may rewrite P' as $P' = P[\alpha] + P[\alpha, \beta]S'P[\beta, \alpha]$.

3. [BLM05, Theorem 1.23] Assume that the Markov chain is irreducible and positive recurrent. Partition the stationary distribution π as $\pi = [\pi_\alpha, \pi_\beta]$, with $\pi_\alpha = (\pi_i)_{i \in \alpha}$ and $\pi_\beta = (\pi_i)_{i \in \beta}$. Then one has $\pi_\alpha P' = \pi_\alpha$, $\pi_\beta = \pi_\alpha P[\alpha, \beta]S'$, where P' and S' are defined in Fact 2.
4. [BLM05, Secs. 1.6, 3.5, 4.5] In the case of finite Markov chains, censoring is equivalent to Schur complementation (see Fact 6 of Section 54.4). Censoring is at the basis of several numerical methods for computing the invariant probability vector π ; indeed, from Fact 3, if the invariant probability vector π_α associated with the censored process is available, the vector π_β can be easily computed from π_α . A smart choice of the sets α and β can lead to an efficient computation of the stationary distribution. As an example of this fact, Ramaswami's recursive formula for computing the vector π associated with an M/G/1-type Markov chain is based on successive censorings, where the set A is finite, and $\beta = \mathbb{N} - \alpha$.

54.6 Numerical Methods

Definitions:

Let A be a $k \times k$ matrix. A splitting $A = M - N$, where $\det M \neq 0$, is a **regular splitting** if $M^{-1} \geq 0$ and $N \geq 0$. A regular splitting is said to be **semiconvergent** if the matrix $M^{-1}N$ is semiconvergent.

Facts:

1. The main computational problem in Markov chains is the computation of the invariant probability vector π . If the space state E is finite and the Markov chain is irreducible, classical techniques for solving linear systems can be adapted and specialized to this purpose. We refer the reader to the book [Ste94] for a comprehensive treatment on these numerical methods; in facts below we analyze the methods based on LU factorization and on regular splittings of M_0 -matrices. If the space state E is infinite, general numerical methods for computing π can be hardly designed. Usually, when the state space is infinite, the matrix P has some structures which are specific of the real problem modeled by the Markov chain, and numerical methods for computing π which exploit the structure of P can be designed. Quite common structures arising in queueing problems are the (block) tridiagonal structure, the (block) Hessenberg structure, and the (block) Toeplitz structure (see, for instance, Example 4 Section 54.1). We refer the reader to the book [BLM05] for a treatment on the numerical solution of Markov chains, where the matrix P is infinite and structured.

2. [BP94, Cor. (4.17), Chap. 6], [Ste94, Sec. 2.3] If P is a $k \times k$ irreducible stochastic matrix, then the matrix $I - P$ has a unique LU factorization $I - P = LU$, where L is a lower triangular matrix with unit diagonal entries and U is upper triangular. Moreover, L and U are M_0 -matrices, U is singular, $U\mathbf{1}_k = 0$, and $(U)_{k,k} = 0$.
3. [Ste94, Sec. 2.5] The computation of the LU factorization of an M_0 -matrix by means of Gaussian elimination involves additions of nonnegative numbers in the computation of the off-diagonal entries of L and U , and subtractions of nonnegative numbers in the computation of the diagonal entries of U . Therefore, numerical cancellation cannot occur in the computation of the off-diagonal entries of L and U . In order to avoid possible cancellation errors in computing the diagonal entries of U , Grassmann, Taksar, and Heyman have introduced in [GTH85] a simple trick, which fully exploits the property that U is an M_0 -matrix such that $U\mathbf{1}_k = 0$. At the general step of the elimination procedure, the diagonal entries are not updated by means of the classical elimination formulas, but are computed as minus the sum of the off-diagonal entries. Details are given in Algorithm 1.
4. If $I - P = LU$ is the LU factorization of $I - P$, the invariant probability vector π is computed by solving the two triangular systems $\mathbf{y}L = \mathbf{x}$ and $\mathbf{x}U = 0$, where \mathbf{x} and \mathbf{y} are row vectors, and then by normalizing the solution \mathbf{y} . From Fact 2, the nontrivial solution of the latter system is $\mathbf{x} = \alpha \mathbf{e}_k^T$, for any $\alpha \neq 0$. Therefore, if we choose $\alpha = 1$, the vector π can be simply computed by solving the system $\mathbf{y}L = \mathbf{e}_k^T$ by means of back substitution, and then by setting $\pi = (1/\mathbf{y}\mathbf{1})\mathbf{y}$, as described in Algorithm 2.
5. [BP94, Cor. (4.17), Chap. 6] If P is a $k \times k$ irreducible stochastic matrix, the matrix $I - P^T$ has a unique LU factorization $I - P^T = LU$, where L is a lower triangular matrix with unit diagonal entries, and U is upper triangular. Moreover, both L and U are M_0 -matrices, U is singular, and $(U)_{k,k} = 0$. Since L is nonsingular, the invariant probability vector can be computed by calculating a nonnegative solution of the system $U\mathbf{y} = 0$ by means of back substitution, and by setting $\pi = (1/\|\mathbf{y}\|_1)\mathbf{y}^T$.
6. If P is a $k \times k$ irreducible stochastic matrix, and if \hat{P} is the $(k-1) \times (k-1)$ matrix obtained by removing the i th row and the i th column of P , where $i \in \{1, \dots, k\}$, then the matrices $I - \hat{P}^T$ and $I - \hat{P}$ have a unique LU factorization, where both the factors L and U are M -matrices. Therefore, if $i = k$ and if the matrix $I - P^T$ is partitioned as

$$I - P^T = \begin{bmatrix} I - \hat{P}^T & \mathbf{a} \\ \mathbf{b}^T & c \end{bmatrix},$$

one finds that the vector $\hat{\pi} = [\pi_1, \dots, \pi_{k-1}]^T$ solves the nonsingular system $(I - \hat{P}^T)\hat{\pi} = -\pi_k \mathbf{a}$. Therefore, the vector π can be computed by solving the system $(I - \hat{P}^T)\mathbf{y} = -\mathbf{a}$, say by means of LU factorization, and then by setting $\pi = [\mathbf{y}^T, 1]/(1 + \|\mathbf{y}\|_1)$.

7. [Ste94, Sec. 3.6] Let P be a finite irreducible stochastic matrix, and let $I - P^T = M - N$ be a regular splitting of $I - P^T$. Then the matrix $H = M^{-1}N$ has spectral radius 1 and 1 is a simple eigenvalue of H .
8. [Ste94, Theorem 3.3], [BP94], [Var00] Fact 7 above is not sufficient to guarantee that any regular splitting of $I - P^T$, where P is a finite stochastic irreducible matrix, is semiconvergent (see the Example below). In order to be semiconvergent, the matrix H must not have eigenvalues of modulus 1 different from 1, and the Jordan blocks associated with 1 must have dimension 1.
9. An iterative method for computing π based on a regular splitting $I - P^T = M - N$ consists in generating the sequence $\mathbf{x}_{n+1} = M^{-1}N\mathbf{x}_n$, for $n \geq 0$, starting from an initial column vector \mathbf{x}_0 . If the regular splitting is semiconvergent, by choosing $\mathbf{x}_0 \geq 0$ such that $\|\mathbf{x}_0\|_1 = 1$, the sequence $\{\mathbf{x}_n\}_n$ converges to π^T . The convergence is linear and the asymptotic rate of convergence is the second largest modulus eigenvalue θ of

$M^{-1}N$.

10. [Ste94, Theorem 3.6] If P is a finite stochastic matrix, the methods of Gauss–Seidel, Jacobi, and SOR, for $0 < \omega \leq 1$, applied to $I - P^T$, are based on regular splittings.
11. If E is finite, the invariant probability vector is, up to a multiplicative constant, the nonnegative left eigenvector associated with the dominating eigenvalue of P . The power method applied to the matrix P^T is the iterative method defined by the trivial regular splitting $I - P^T = M - N$, where $M = I$, $N = P^T$. If P is primitive, the power method is convergent.
12. [Ste94, Theorems 3.6, 3.7] If P is a $k \times k$ irreducible stochastic matrix, and if \hat{P} is the $(k-1) \times (k-1)$ matrix obtained by removing the i th row and the i th column of P , where $i \in \{1, \dots, k\}$, then any regular splitting of the matrix $I - \hat{P}^T = M - N$ is convergent. In particular, the methods of Gauss–Seidel, Jacobi, and SOR, for $0 < \omega \leq 1$, applied to $I - \hat{P}^T$, being based on regular splittings, are convergent.
13. [Ste94, Theorem 3.17] Let P be a $k \times k$ irreducible stochastic matrix and let $\epsilon > 0$. Then the splitting $I - P^T = M_\epsilon - N_\epsilon$, where $M_\epsilon = D - L + \epsilon I$ and $N_\epsilon = U + \epsilon I$, is a semiconvergent regular splitting for every $\epsilon > 0$.
14. [Ste94, Theorem 3.18] Let P be a $k \times k$ irreducible stochastic matrix, let $I - P^T = M - N$ be a regular splitting of $I - P^T$, and let $H = M^{-1}N$. Then the matrix $H_\alpha = (1 - \alpha)I + \alpha H$ is semiconvergent for any $0 < \alpha < 1$.

Algorithms:

1. The following algorithm computes the LU factorization of $I - P$ by using the Grassmann, Taksar, Heyman (GTH) trick of Fact 3:

Computation of the LU factorization of $I - P$ with GTH trick

Input: the $k \times k$ irreducible stochastic matrix P .

Output: the matrix $A = [a_{i,j}]$ such that $a_{i,j} = l_{i,j}$ for $i > j$, and $a_{i,j} = u_{i,j}$ for $i \leq j$, where $L = [l_{i,j}]$, $U = [u_{i,j}]$ are the factors of the LU factorization of $I - P$.

Computation:

- 1– set $A = I - P$;
- 2– for $m = 1, \dots, k-1$:
 - (a) for $i = m+1, \dots, k$, set $a_{i,m} = a_{i,m}/a_{m,m}$;
 - (b) for $i, j = m+1, \dots, k$, set $a_{i,j} = a_{i,j} - a_{i,m}a_{m,j}$ if $i \neq j$;
 - (c) for $i = m+1, \dots, k$ set $a_{i,i} = -\sum_{l=m+1, l \neq i}^k a_{l,i}$.

2. The algorithm derived by Fact 4 above is the following:

Computation of π through LU factorization

Input: the factor $L = [l_{i,j}]$ of the LU factorization of the matrix $I - P$, where P is a $k \times k$ irreducible stochastic matrix.

Output: the invariant probability vector $\pi = (\pi_i)_{i=1,k}$.

Computation:

- 1– set $\pi_k = 1$;
- 2– for $j = k-1, \dots, 1$ set $\pi_j = -\sum_{i=j+1}^k \pi_i l_{i,j}$;
- 3– set $\pi = (\sum_{i=1}^k \pi_i)^{-1} \pi$.

Examples:

The matrix

$$P = \begin{bmatrix} 0.5 & 0 & 0 & 0.5 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0.1 & 0.9 \end{bmatrix}$$

is the transition matrix of an irreducible and aperiodic Markov chain. The splitting $I - P^T = M - N$ associated with the Jacobi method is a regular splitting. One may easily verify that the iteration matrix is

$$M^{-1}N = D \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} D^{-1},$$

where $D = \text{diag}(1, \frac{5}{2}, \frac{5}{6}, 5)$. Therefore, the regular splitting is not semiconvergent since the iteration matrix has period 4 (see Fact 8 above). Also the splitting $I - P^T = M - N$ associated with Gauss-Seidel method is a regular splitting; however, in this case the iteration matrix is

$$M^{-1}N = \begin{bmatrix} 0 & \frac{2}{5} & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & \frac{2}{3} \\ 0 & 2 & 0 & 0 \end{bmatrix},$$

which is semiconvergent.

References

- [ANT02] A. Arasu, J. Novak, A. Tomkins, and J. Tomlin. PageRank computation and the structure of the Web. <http://www2002.org/CDROM/poster/173.pdf>, 2002.
- [BLM05] D.A. Bini, G. Latouche, and B. Meini. *Numerical Methods for Structured Markov Chains*. Series on Numerical Mathematics and Scientific Computation. Oxford University Press Inc., New York, 2005.
- [BP94] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, Vol. 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994. (Revised reprint of the 1979 original.)
- [Çin75] E. Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, Upper Saddle River, N.J., 1975.
- [GTH85] W.K. Grassmann, M.I. Taksar, and D.P. Heyman. Regenerative analysis and steady state distributions for Markov chains. *Oper. Res.*, 33(5):1107–1116, 1985.
- [LR99] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1999.
- [Mey89] C.D. Meyer. Stochastic complementation, uncoupling Markov chains, and theory of nearly reducible systems. *SIAM Rev.*, 31(2):240–272, 1989.
- [Mol04] C. Moler. *Numerical computing with MATLAB*. <http://www.mathworks.com/moler>, 2004. (Electronic edition.)
- [Neu89] M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Vol. 5 of *Probability: Pure and Applied*. Marcel Dekker, New York, 1989.
- [Nor99] J.R. Norris. *Markov Chains*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, U.K., 3rd ed., 1999.

- [PBM99] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the Web, 1999. <http://dbpubs.stanford.edu:8090/pub/1999-66>.
- [Ste94] W.J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, NJ, 1994.
- [Var00] R.S. Varga. *Matrix Iterative Analysis*, Vol. 27 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, expanded edition, 2000.

