Markov chains

Métodos Estocásticos de IO
MBA mención en Operaciones

Mario González
Facultad de Ciencias Administrativas
Escuela Politécnica Nacional

July 5, 2017

Learning Objectives

- Introduce Markov chains and their properties

Markov chains I

- Markov chains were first introduced in 1906 by Andrey Markov.
- Let us start by considering an i.i.d. sequence of random variables. $X_0, X_1, \ldots, X_n, \ldots$
- Independence can be an excessively restrictive assumption; it means that the X_n provide absolutely no information about each other.
- At the other extreme, allowing arbitrary interactions between the X_n makes it very difficult to compute even basic things.
- A Markov chain is a sequence of r.v.s that exhibits one-step dependence.
- Markov chains are a happy medium between complete independence and complete dependence.

Markov chains II

- Intuitively, a Markov chain is a discrete parameter process in which the future is independent of the past given the present. For example, suppose that we decided to play a game with a fair, unbiased coin. We each start with five nickles and repeatedly toss the coin. If it turns up heads, then you give me a nickle; if tails, I give you a nickle. We continue until one of us has none and the other has ten nickles. The sequence of heads and tails from the successive tosses of the coin would form an i.i.d. stochastic process; the sequence representing the total number of nickles that you hold would be a Markov chain. To see this, assume that after several tosses, you currently have three nickles. The probability that after the next toss you will have four nickles is 0.5 and knowledge of the past (i.e., how many nickles you had one or two tosses ago) does not help in calculating the probability of 0.5. Thus, the future (how many nickles you will have after the next toss) is independent of the past (how many nickles you had several tosses ago) given the present (you currently have three nickles).

Markov property and transition matrix I

- Markov chains “live” in both space and time: the set of possible values of the X_n is called the **state space**, and the index n represents the evolution of the process over **time**.
- The state space of a Markov chain can be either discrete or continuous, and time can also be either discrete or continuous.
- We will focus exclusively on discrete-state, discrete-time Markov chains, with a finite state space.
- Specifically, we will assume that the X_n take values in a finite set, which we usually take to be $\{1, 2, \dots, M\}$ or $\{0, 1, \dots, M\}$.

Markov property and transition matrix II

Markov chain

A sequence of random variables $X_0, X_1, X_2, \ldots$ taking values in the state space $\{1, 2, \ldots, M\}$ is called a Markov chain if for all $n \geq 0$,

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

- The quantity $P(X_{n+1} = j | X_n = i)$ is called the **transition probability** from state i to state j.
- We will implicitly assume that a Markov chain is **time-homogeneous**, which means that the transition probability $P(X_{n+1} = j | X_n = i)$ is the same for all times n.
- The condition above is called the **Markov property**, and it says that given the entire past history $X_0, X_1, X_2, \ldots, X_n$, only the most recent term, X_n, matters for predicting X_{n+1}.

Markov property and transition matrix III

Transition Matrix

Let $X_0, X_1, X_2, \ldots$ be a Markov chain with state space $\{1, 2, \ldots, M\}$ and let $q_{ij} = P(X_{n+1} = j | X_n = i)$ be the transition probability from state i to state j. The $M \times M$ matrix $Q = (q_{ij})$ is called the **transition matrix** of the chain.

- Note that Q is a nonnegative matrix in which each row sums to 1. This is because, starting from any state i, the events move to 1, move to 2, ... , move to M are disjoint, and their union has probability 1 because the chain has to go somewhere.

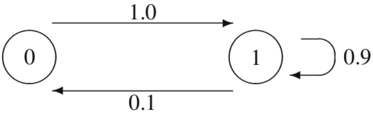
MC Examples I

- Consider a farmer using an old tractor. The tractor is often in the repair shop but it always takes only one day to get it running again. The first day out of the shop it always works but on any given day thereafter, independent of its previous history, there is a 10% chance of it not working and thus being sent back to the shop. Let $X_0, X_1, \ldots$ be random variables denoting the daily condition of the tractor, where a one denotes the working condition and a zero denotes the failed condition. In other words, $X_n = 1$ denotes that the tractor is working on day n and $X_n = 0$ denotes it being in the repair shop on day n. Thus, $X_0, X_1, \ldots$ is a Markov chain with state space $E = \{0, 1\}$ and with transition matrix

$$Q = \begin{bmatrix} 0 & 1 \\ 0.1 & 0.9 \end{bmatrix}$$

MC Examples II

- In order to develop a mental image of the Markov chain, it is very helpful to draw a state diagram of the transition matrix.
- In the diagram, each state is represented by a circle and the transitions with positive probabilities are represented by an arrow:



MC Examples III

- A salesman lives in town ‘a’ and is responsible for towns ‘a’, ‘b’, and ‘c’. Each week he is required to visit a different town. When he is in his home town, it makes no difference which town he visits next so he flips a coin and if it is heads he goes to ‘b’ and if tails he goes to ‘c’. However, after spending a week away from home he has a slight preference for going home so when he is in either towns ‘b’ or ‘c’ he flips two coins. If two heads occur, then he goes to the other town; otherwise he goes to ‘a’. The successive towns that he visits form a Markov chain with state space $E = \{a, b, c\}$ where the random variable X_n equals a, b, or c according to his location during week n.
  - Build the state diagram for this system and the associated Markov matrix.

MC Examples IV

- It is often true that the “steps” of a Markov chain refer to days, weeks, or months, but that need not be the case. Consider a page of text and represent vowels by zeros and consonants by ones. Thus the page becomes a string of zeros and ones. It has been indicated that the sequence of vowels and consonants in the Samoan language forms a Markov chain, where a vowel always follows a consonant and a vowel follows another vowel with a probability of 0.51.
    - Build the state diagram for this system and the associated Markov matrix.

Multi-step transitions I

- The Markov matrix provides direct information about one-step transition probabilities and it can also be used in the calculation of probabilities for transitions involving more than one step. Consider the salesman Example starting in town b. The Markov matrix indicates that the probability of being in State a after one step (in one week) is 0.75, but what is the probability that he will be in State a after two steps?

$$
\begin{aligned}
Pr\{X_2 = a|X_0 = b\} = \quad & Pr\{X_1 = a|X_0 = b\} \times Pr\{X_2 = a|X_1 = a\} \\
& + Pr\{X_1 = b|X0 = b\} \times Pr\{X_2 = a|X_1 = b\} \\
& + Pr\{X_1 = c|X0 = b\} \times Pr\{X_2 = a|X_1 = c\} \\
= \quad & Q(b,a)Q(a,a) + Q(b,b)Q(b,a) + Q(b,c)Q(c,a).
\end{aligned}
$$

- The final equation should be recognized as the definition of matrix multiplication; thus, $Pr\{X_2 = a|X_0 = b\} = Q^2(b,a)$.

Multi-step transitions II

n-step transition probability

The n-step transition probability from i to j is the probability of being at j exactly n steps after being at i. Wedenote this by $q_{ij}^{(n)}$:

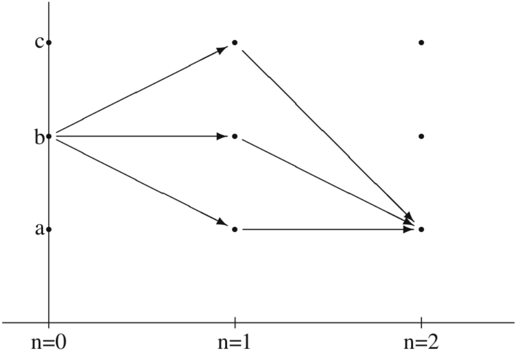
$$q_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

The nth power of the transition matrix gives the n-step transition probabilities

$$q_{ij}^{(n)} \text{ is the } (i,j) \text{ entry of } Q^n.$$

Multi-step transitions III

- Possible paths of a two-step transition from State b to State a for a three-state Markov chain (Travel salesman example).



Multi-step transitions IV

- Markov chains are often used to analyze the cost or profit of an operation and thus we need to consider a cost or profit function imposed on the process. For example, suppose in Travel salesman Example that every week spent in town a resulted in a profit of \$1000, every week spent in town b resulted in a profit of \$1200, and every week spent in town c resulted in a profit of \$1250. We then might ask what would be the expected profit after the first week if the initial town was town a? Or, more generally, what would be the expected profit of the nth week if the initial town was a?

Multi-step transitions V

Property

Let $X = X_n; n = 0, 1, \ldots$ be a Markov chain with state space E, Markov matrix Q, and profit function f (i.e., each time the chain visits state i, a profit of $f(i)$ is obtained). The expected profit at the nth step is given by

$$E[f(X_n)|X_0 = i] = \mathbf{Q^n f}(i).$$

- Getting back to our TS example

$$E[f(X_2)|X_0 = a] = 0.75 \times 1000 + 0.125 \times 1200 + 0.125 \times 1250 = 1056.25.$$

Multi-step transitions VI

- Up until now we have always assumed that the initial state was known. However, that is not always the situation. The manager of the traveling salesman might not know for sure the location of the salesman; instead, all that is known is a probability mass function describing his initial location. We now ask, what is the probability that the salesman will be in town a next week given an initial probability vector.

Property

Let $X = X_n; n = 0, 1, \ldots$ be a Markov chain with state space E, Markov matrix Q, and initial probability vector $\boldsymbol{\mu}$ (i.e., $\mu(i) = Pr X_0 = i$). Then

$$Pr_{\mu}\{X_n = j\} = \boldsymbol{\mu}\mathbf{Q}^n(j).$$

Multi-step transitions VII

- The last two properties can be combined as:

Property

Let $X = X_n; n = 0, 1, \ldots$ be a Markov chain with state space E, Markov matrix Q, and initial probability vector $\boldsymbol{\mu}$ and profit function $\mathbf{f}$. The expected profit at the nth step is given by

$$Pr_\mu\{X_n = j\} = \boldsymbol{\mu}\mathbf{Q}^n\mathbf{f}.$$

Multi-step transitions VIII

- For our TS example:

$$
\begin{aligned}
\boldsymbol{\mu}\mathbf{P}^2\mathbf{f} &= (0.50, 0.30, 0.20) \begin{bmatrix} 0.75 & 0.125 & 0.125 \\ 0.1875 & 0.4375 & 0.375 \\ 0.1875 & 0.375 & 0.4375 \end{bmatrix} \begin{pmatrix} 1000 \\ 1200 \\ 1250 \end{pmatrix} \\
&= 1119.375 \,.
\end{aligned}
$$

Multi-step transitions IX

- Example. Given the Markov matrix:

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

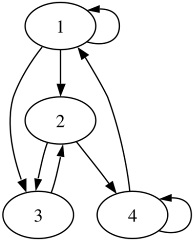
  Suppose the initial conditions are $\mu = (1/4, 1/4, 1/4, 1/4)$, meaning that the chain has equal probability of starting in each of the four states. Calculate the marginal distribution of X_5.

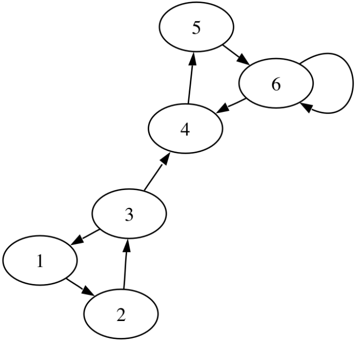
Classification of states I

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

Classification of states II

- 4-state Markov chain with all states recurrent. Right: 6-state Markov chain with states 1, 2, and 3 transient.





Classification of states III

Recurrent and transient states

State i of a Markov chain is **recurrent** if starting from i, the probability is 1 that the chain will eventually return to i. Otherwise, the state is **transient**, which means that if the chain starts from i, there is a positive probability of never returning to i.

Classification of states IV

- Example. Given the Markov matrix:

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.2 & 0 & 0.1 & 0.7 \end{bmatrix}$$

and diagram

Classification of states V

- The first passage probabilty $F(i,j)$, which is the probability of eventually reaching State j at least once given that the initial state was i. The probability $F(i,j)$ is defined by

$$F(i,j) = Pr\{T_j < \infty | X_0 = i\}$$

- By inspecting the state diagram in the previous example, it should be obvious that the first passage probabilities are given by

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ < 1 & < 1 & < 1 & < 1 \end{bmatrix}$$

Classification of states VI

- The expected number of visits to State j given the initial state was i is denoted by $R(i, j)$ ("R" for returns) and is defined by

$$R(i, j) = E[N_j | X_0 = i].$$

Again the state diagram of our example allows the determination of some of the values of R as follows:

$$R = \begin{bmatrix} \infty & 0 & 0 & 0 \\ 0 & \infty & \infty & 0 \\ 0 & \infty & \infty & 0 \\ \infty & \infty & \infty & < \infty \end{bmatrix}$$

- A state j is called **transient** if $F(j, j) < 1$. Equivalently, state j is transient if $R(j, j) < \infty$.
- A state j is called **recurrent** if $F(j, j) = 1$. Equivalently, state j is recurrent if $R(j, j) = \infty$.
- It is only the diagonal elements of the F and R matrices that determine whether a state is transient or recurrent.

Classification of states VII

- Example. Given the Markov matrix:

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

  Classify each state.

Classification of states VIII

- Along with classifying states, we also need to be able to classify sets of states.
- We will first define a **closed** set which is a set that once the Markov chain has entered the set, it cannot leave the set.

Definition

Let $X = \{Xn; n = 0, 1, \dots\}$ be a Markov chain with Markov matrix Q and let C be a set of states contained in its state space. Then C is closed if

$$\sum_{j \in C} Q(i, j) = 1, \forall i \in C.$$

Classification of states IX

- To illustrate the concept of **closed sets**, let us refer to the state transition diagram below.

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.3 & 0.7 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.2 & 0 & 0.1 & 0.7 \end{bmatrix}$$

  - The sets $\{2, 3, 4\}$, $\{2\}$, and $\{3, 4\}$ are not closed sets. The set $\{1, 2, 3, 4\}$ is obviously closed, but it can be reduced to a smaller closed set. The set $\{1, 2, 3\}$ is also closed, but again it can be further reduced. Both sets $\{1\}$ and $\{2, 3\}$ are closed and cannot be reduced further. This idea of taking a closed set and trying to reduce it is extremely important and leads to the definition of an **irreducible set**.

Classification of states X

Definition

A closed set of states that contains no proper subset which is also closed is called **irreducible**. A state that forms an irreducible set by itself is called an **absorbing state**.

- The Markov chain of the Example before has two irreducible sets: the set $\{1\}$ and the set $\{2, 3\}$. Since the first irreducible set contains only one state, that state is an absorbing state.

Classification of states XI

- Example. Let X be a Markov chain with state space $E = \{a, \cdots, g\}$ and Markov matrix

$$Q = \begin{bmatrix} 0.3 & 0.7 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.4 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & 0.1 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.7 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 \end{bmatrix}$$

- Draw the state transition diagram identify the closed and irreducible sets.

Classification of states XII

Definition

Let C be an irreducible set of states such that the number of states within C is finite. Then each state within C is recurrent.

Definition

The closed set of states C is irreducible if and only if every state within C communicates with every other state within C.

- Two states, i and j, communicate if and only if it is possible to eventually reach j from i and it is possible to eventually reach i from j.
- The communication must be both ways but it does not have to be in one step.

Classification of states XIII

- Example. Let X be a Markov chain with state space $E = \{a, \cdots, g\}$ and Markov matrix

$$Q = \begin{bmatrix} 0.3 & 0.1 & 0.2 & 0.2 & 0.1 & 0.1 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.2 & 0 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0.3 & 0.4 & 0.1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 & 0 & 0 & 0.2 \end{bmatrix}$$

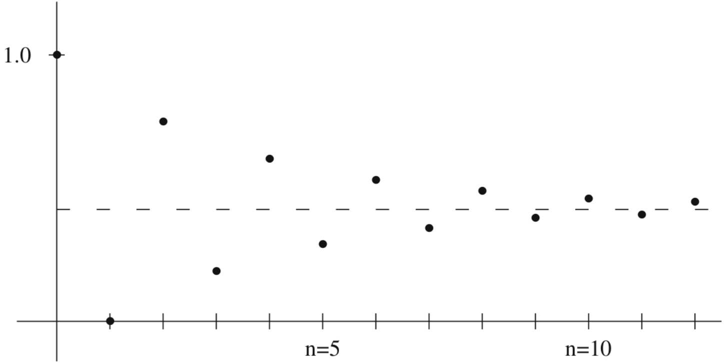
- Draw the state diagram find the irreducible recurrent sets and transient states.

Steady-State Behavior I

- Let us get back to our TS example.
- A salesman lives in town 'a' and is responsible for towns 'a', 'b', and 'c'. Each week he is required to visit a different town. When he is in his home town, it makes no difference which town he visits next so he flips a coin and if it is heads he goes to 'b' and if tails he goes to 'c'. However, after spending a week away from home he has a slight preference for going home so when he is in either towns 'b' or 'c' he flips two coins. If two heads occur, then he goes to the other town; otherwise he goes to 'a'. The successive towns that he visits form a Markov chain with state space $E = \{a, b, c\}$ where the random variable X_n equals a, b, or c according to his location during week n.

Steady-State Behavior II

- Values of $P_n(a, a)$ as a function of n.



- One can appreciate that $\lim_{n\to\infty} PrX_n = a|X_0 = a = 0.42857$.
- If you spent the time to graph the probabilities of being in State a starting from State b instead of State a, you would discover the same limiting value.

Steady-State Behavior III

- When discussing steady-state (or limiting) conditions, this is what is meant, not that the chain stops changing.
- It is dynamic by definition, but that after enough time has elapsed the probabilities do not change with respect to time.
- It is often stated that steady-state results are independent of initial conditions.
- If the entire state space of a Markov chain forms an irreducible recurrent set, the Markov chain is called an **irreducible recurrent Markov chain**.
- In this case the steady-state probabilities are independent of the initial state and are not difficult to compute as is seen in the following property.

Steady-State Behavior IV

Property

Let $X = \{X_n; n = 0, 1, \cdots\}$ be a Markov chain with finite state space E and Markov matrix Q. Furthermore, assume that the entire state space forms an irreducible, recurrent set, and let

$$\pi(j) = \lim_{n\to\infty} Pr\{X_n = j | X_0 = i\}.$$

The vector $\boldsymbol{\pi}$ is the solution to the following system

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi},$$
$$\sum_{i\in E} \pi(i) = 1.$$

Steady-State Behavior V

- To illustrate the determination of $\boldsymbol{\pi}$, observe that the Markov chain of TS Example is irreducible, recurrent. Applying the Property before, we obtain

$$
\begin{aligned}
0.75\pi_b + 0.75\pi_c &= \pi_a \\
0.5\pi_a \qquad\qquad + 0.25\pi_c &= \pi_b \\
0.5\pi_a + 0.25\pi_b \qquad\qquad &= \pi_c \\
\pi_a + \pi_b + \pi_c &= 1 \,.
\end{aligned}
$$

- There are four equations and only three variables, so normally there would not be a unique solution; however, for an irreducible Markov matrix there is always exactly one redundant equation from the system formed by $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. Thus, to solve the above system, arbitrarily choose one of the first three equations to discard and solve the remaining 3 by 3 system (never discard the final or norming equation) which yields

$$\pi_a = \frac{3}{7}, \pi_b = \frac{2}{7}, \pi_c = \frac{2}{7}.$$

Sources and resources

- Introduction to Probability, Joseph K. Blitzstein, Jessica Hwang.
- Applied Probability and Stochastic Processes, Feldman, Richard M., Valdez-Flores, Ciriaco.