Discrete Markov Chain. Theory and use

Andres Vallone
PhD Student
andres.vallone@predoc.uam.es

ECONRES

2016
Contents

1 Introduction

2 Discrete Markov Chains (DMC)
   - Concept and definition
   - Examples
   - Transitions Matrix
   - Chains Classification

3 Empirical Uses
   - Introduction
   - Test the Markov Property
   - Estimating the Transition Matrix
   - Example
Introduction
What is the probability of having a nice day tomorrow if today is raining?

What is the probability of buying a Coke when you first buy a Pepsi?

What is the probability of a state moving out of poverty?

How was the evolution of the income distribution in a country?
What is the probability of have a nice day tomorrow if today is raining?

What is the probability of buy a Coke when you first buy a Pepsi?

What is the probability of a state move out of poverty?

How was the evolution of the income distribution in a country?

All of this questions can be answering by the use of a Discrete Markov Chain.
Discrete Markov Chains (DMC)
What is a DMC?

- The Markov Chain is a discrete-time stochastic process
- Describes a system whose states change over time
- Changes are governed by a probability distribution
- Is a "memorylessness" process: the next state only depends upon the current system state, the path to the present state is not relevant
- In order to make a formal definition suppose that exists:
  - A set of independent state $S = \{s_1, \ldots, s_k\}$
  - A sequence of random variables $X_0, X_1, X_2, \ldots$ with values in $S$ that describes the state of the system in time $t$
  - A set of transition rules specifying the probability of move from the state $i$ in time $t$ to the state $j$ in time $t + 1$
- If the rules are independent of time, the DMC is homogeneous.
Discrete Markov Chain

Definition

A random process $X_0, X_1, X_2, \ldots$ with finite state space $S = \{s_1, \ldots, s_k\}$ is said to be a Markov homogeneous chain if $\forall i, j \in \{i, \ldots, k\}$ and $\forall i_0, \ldots, i_{n-1} \in \{1, \ldots, k\}$

$$P[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i_{n-1}, X_t = i] = P[X_{t+1} = j \mid X_t = i] = p_{i,j}$$

- $p_{i,j}$ is a conditional probability: defines the probability that the chain jumps to state $j$, at time $t + 1$, given that it is in state $i$ at time $t$,

- $p_{i,j}$ is a rule of movement, since $S$ has size $k$, exist $k \times k$ rules of movement. The set of rules specify the transition matrix.
The Transition Matrix

\[ P = \begin{bmatrix}
  p_{11} & p_{12} & \cdots & p_{1k} \\
  p_{21} & p_{22} & \cdots & p_{2k} \\
  \vdots & \vdots & \ddots & \vdots \\
  p_{k1} & p_{k2} & \cdots & p_{kk}
\end{bmatrix} \]

- Each \( p_{ij} \) denote the probability of one step in time.
- Are probabilities \( \forall p_{i,j} \geq 0 \)
- All state are exclusive \( \sum_{i=1}^{k} p_{ij} = 1 \)
- In case of an homogeneous DMC \( p_{ij} = p_{i,j} \ \forall \ t \)
- Homogeneity implies that every time the chain is in state \( s \), the probability of jumping to another state is the same.
According to Kemeny, Snell, and Thompson\(^1\), Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day.

- What are the transition matrix in this case?

Example 1

Transition Matrix

\[ P = \begin{bmatrix}
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2}
\end{bmatrix} \]
Coke or Pepsi?

- Given that a person’s last cola purchase was Coke, there is a 90% chance that his next cola purchase will also be Coke.
- If a person’s last cola purchase was Pepsi, there is an 80% chance that his next cola purchase will also be Pepsi.
- What are the transition matrix in this case?
Example 2

Transition Matrix

\[ P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \]
The Transition Matrix

Consider the example 1, what is the probability of have an snowy day in two days, if today is raining?

- Have a snowy day in two days denote as $p_{13}^{(2)}$ is the disjoint union of:
  - (a) it is rainy tomorrow and snowy two days from now
  - (b) it is nice tomorrow and snowy two days from now
  - (c) it is snowy tomorrow and snowy two days from now

$$p_{13}^{(2)} = p_{11}p_{13} + p_{12}p_{21} + p_{13}p_{33}$$

- This is just what is done in obtaining the 1,3-entry of the product of the P with itself.
The Transition Matrix

Proposition (Chapman-Kolgomorov equation)

For any $n > 0, m > 0$, $i \in S$, $j \in S$

$$p_{ij}^{m+n} = \sum_{k \in S} p_{i,k}^n p_{jk}^m$$
The Transition Matrix

Proposition (Chapman-Kolgomorov equation)

For any $n > 0, m > 0, i \in S, j \in S$

$$p_{ij}^{m+n} = \sum_{k \in S} p_{i,k}^n p_{jk}^m$$

Theorem

Let $P$ be the transition matrix of a Markov chain. The $ij$-th entry of the matrix $P^n$ gives the probability that the Markov chain, starting in state $s_i$, will be in state $s_j$ after $n$ steps.
Long term behaviour

Let \( \pi_0 \) a vector \( k \times 1 \) defining for all state \( i \) the probability that the Markov chain is initially at \( i \)

\[ \pi_0 (i) = P [X_0 = i] : i = 1, \ldots, k \]

Consider a vector \( \pi_n (j) \) \( k \times 1 \) that defines the probabilities that the Markov chain is at state \( j \) after \( n \) step

\[ \pi_n (j) = \{ \pi_n (1), \ldots, \pi_n (k) \} \]

Theorem

Let \( P \) be the transition matrix of a Markov chain, and let \( \pi_0 \) be the probability vector which represents the starting distribution. Then the probability that the chain is in state \( s_i \) after \( n \) steps is the \( i \)th entry of \( \pi_n = P^n \pi_0 \)
Chains Classification

Periodic State

This means at every state \(i\), if \(n\) is an even number \(P^n(i, i) > 0\)
Chains Classification

Absorbing State

\[ P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0 & 0.5 & 0 & 0 & 0 \\
0 & 0.5 & 0 & 0.5 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \]

- rows are not identical, i.e., the chain does not forget the initial state
Chains Classification

Absorbing State

\[
P^n = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0.75 & 0 & 0 & 0 & 0.25 \\
0.5 & 0 & 0 & 0 & 0.5 \\
0.25 & 0 & 0 & 0 & 0.75 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

- states 2, 3, 4 are transient: \(p^n_{i2} = p^n_{i3} = p^n_{i4} = 0\) i.e. after passing a large amount of time the chain will stop visiting these states.
Chains Classification

Reducible Chains

\[ P = \begin{bmatrix} 0.2 & 0.8 & 0 & 0. & 0 & 0.2 & 0.1 & 0.9 & 0 & 0. & 0. & 0 \\ 0.1 & 0.9 & 0 & 0 & 0 & 0.3 & 0.7 & 0 & 0. & 0.2 & 0.6 & 0.2 \\ 0 & 0 & 0 & 0.7 & 0 & 0 & 0.7 & 0 & 0.2 & 0.6 & 0.2 & 0.3 \\ \end{bmatrix} \]

\[ P^n = \begin{bmatrix} 0.1 & 0.9 & 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 & 0.1 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & 0.2 & 0.6 & 0.2 \\ \end{bmatrix} \]
Chains Classification

Ergodic Chains

- State $j$ is reachable from state $i$ if probability to go from $i$ to $j$ in $n \geq 0$ steps is greater then 0
- A subset $X$ of the state space $S$ is closed if $p_{ij} = 0$, $\forall \ i \in S$ and $\forall \ j \in S$
- A closed set of states is irreducible if any state $j \in S$ is reachable from any state $i \in S$
- A Markov Chain is irreducible if the state space $S$ is irreducible
Chains Classification

Irreducible Chains

- Ergodic

- Reducible
Theorem (Basic Limit Theorem)

Any ergodic, aperiodic, Markov chain defined on a set of state $S$ and with a stochastic transition matrix $P$ has a unique stationary distribution $\pi$ with all its components positive. Furthermore, let $P^n$ be the $n$-th power of $P$, then

$$\lim_{n \to \infty} \pi(j) P^n = \pi$$
Empirical Uses
In social science is not common work with discrete state variables.

Wrong discretization produce wrong result by (cite)

In necessary test if the random variables follow a Markov process

Is necessary find the Transition matrix
First, homogeneity over time (time-stationarity) can be checked by dividing the entire sample into $T$ periods, and testing whether or not the transition matrices estimated from each of the $T$ sub-samples differ significantly from the matrix estimated from the entire sample.

\[
\begin{align*}
\{ & H_0 : \forall t \ p_{ij}(t) = p_{ij} \quad t = 1, \ldots, T \\
& H_1 : \forall t \ p_{ij}(t)_{ij} \}
\end{align*}
\]

\[
Q^{(T)} = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \in B_i} n_i(t) \frac{(\hat{p}_{ij}(t) - \hat{p}_{ij})^2}{\hat{p}_{ij}} \sim \text{asy} \chi^2 \left( \sum_{i=1}^{N} (a_i - 1) (b_1 - 1) \right)
\]
Empirical Uses

Test the Markov Property

\[ Q(T) = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \in B_i} n_i(t) \left( \frac{\hat{p}_{ij}(t) - \hat{p}_{ij}}{\hat{p}_{ij}} \right)^2 \sim \text{asy } \chi^2 \left( \sum_{i=1}^{N} (a_i - 1) (b_1 - 1) \right) \]

- \( \hat{p}_{ij} \) denotes the probability of transition from the i-th to the j-th class estimated from the entire sample
- \( \hat{p}_{ij}(t) \) the corresponding transition probability estimated from the t-th sub-sample \( \hat{p}_{ij}(t) = n_{ij}(t) / n_{ij} \)
- \( n_i(t) \) denotes the absolute number of observations initially falling into the i-th class within the t-th sub-sample.
- \( b_i = |B_i| \) is the number of positive entries in the i-th row of the matrix for the entire sample \( B_i = \{ j : \hat{p}_{ij} \} \)
- \( a_1 = |A_i| \) is the number of sub-samples (t) in which observations for the i-th row are available \( A_i = \{ t : n(t) > 0 \} \)
We need to estimate all the $p_{ij}$ parameters of the matrix $k \times k$ of probabilities.

$$p_{ij} = P [X_{t+1} = j \mid X_t = i]$$

We have a sample from the chain $x^n \equiv x_1, \ldots, x_n$ that is a realization of the Markov chain $X_1^n$

$$P [X_1^n = x_1^n] = \prod_{t=1}^{n} P [X_t = x_t \mid X_{t-1} = x_{t-1}]$$

Re-written in term of transitions probability $p_{ij}$

$$L (p) = \prod_{t=2}^{n} p_{x_{t-1}, x_t}$$

Defining $n_{ij}$ as the number of times $i$ is followed by $j$ in $X_1^n$

$$L (p) = \prod_{i=1}^{k} \prod_{j=1}^{k} p_{ij}^{n_{ij}}$$
Estimating the Transition Matrix

\[ L(p) = \prod_{i=1}^{k} \prod_{j=1}^{k} p_{ij}^{n_{ij}} \]

Taking log and remembering that \( \sum_{j} p_{ij} = 1 \) to find \( p_{ij} \) is necessary solve:

\[
\max_{p_{ij}} \ell(p) = \sum_{i,j} n_{ij} \log p_{ij} - \sum_{i=1}^{j} \lambda_i \left( \sum_{j} p_{ij} - 1 \right)
\]

The first order conditions of this problem are:

\[
\frac{\partial \ell}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda_i \equiv 0 \quad (1)
\]

\[
\frac{\partial \ell}{\partial \lambda_i} = \sum_{j} p_{ij} - 1 \equiv 0 \quad (2)
\]
From (1)

\[ p_{ij} = \frac{n_{ij}}{\lambda_i} \quad (3) \]

knowing that we have \( k \) lagrange multipliers and using (2) and (3)

\[
\sum_{j=1}^{k} \frac{n_{ij}}{\lambda_i} \equiv 1
\]

\[
\sum_{j=1}^{k} n_{ij} = \lambda_i \quad (4)
\]

With (3) and (4)

\[ \hat{p}_{ij} = \frac{n_{ij}}{\frac{k}{\sum_{j=1}^{k} n_{ij}}} \]
EN proceso.......