Very short notes on Markov chains

Josep Pijoan-Mas

CEMFI, January 2009

1 Introduction

Definition 1 *A stochastic process* $\{x_t\}$ *is* ***markovian*** *if*

$$Prob\,(x_{t+1}|x_t, x_{t-1}, ..., x_{t-k}) = Prob\,(x_{t+1}|x_t) \quad \forall k \geq 1$$

We say that a given stochastic process displays the *markovian property* or that it is *markovian* when its realization in a given period only depends on the previous period realization and therefore the rest of the history is useless.

A discrete stochastic process $\{x_t\}$ with n possible states displaying the *markovian property* can be characterized by two objects, the $n \times n$ transition matrix Π, which describes the probability of moving from one position in the state space into another, and an $n \times 1$ vector π_t that describes the probability of being in every position of the state space at time t. An element $\Pi_{i,j}$ of the transition matrix Π gives us the probability of moving from position i into position j, that is to say, for any $t > 0$, $\Pi_{i,j} = Prob\,(x_{j,t+1}|x_{i,t})$. Notice that for the markov process to be well specified we need to impose some properties on these objects:

$$\pi_{t,i} \geq 0 \;\; \forall t, i \quad \text{and} \quad \sum_{i=1}^{n} \pi_{t,i} = 1 \tag{1}$$

$$\Pi_{i,j} \geq 0 \;\; \forall j, i \quad \text{and} \quad \sum_{j=1}^{n} \Pi_{i,j} = 1 \quad \forall i = 1, \ldots, n \tag{2}$$

Let's think of $\{x_t\}$ as a vector of individual productivity endowments. Then, the n-dimensional vector π_t describes the state of the system by telling us how many people are of each type at a given period of time t and the transition matrix Π describes the evolution of the system. The evolution of the system is given by:

$$\pi_{t+1} = \Pi^T \pi_t$$

which in less compact notation shows us:

$$\begin{pmatrix} \pi_{1,t+1} \\ \pi_{2,t+1} \\ \dots \\ \pi_{n,t+1} \end{pmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{21} & \dots & \Pi_{n1} \\ \Pi_{12} & \Pi_{22} & \dots & \Pi_{n2} \\ \dots & \dots & \dots & \dots \\ \Pi_{1n} & \Pi_{2n} & \dots & \Pi_{nn} \end{bmatrix} \begin{pmatrix} \pi_{1,t} \\ \pi_{2,t} \\ \dots \\ \pi_{n,t} \end{pmatrix}$$

Notice therefore, that the fraction of people in state i tomorrow is given by the sum of fraction of people in every possible state today times the probability of each of moving towards state i:

$$\pi_{i,t+1} = \sum_{j=1}^{n} \pi_{j,t} \, \Pi_{ji}$$

2 Finding the stationary distribution

Definition 2 *A stationary distribution π^* is one such that:*

$$\pi^* = \Pi^T \pi^* \quad \Leftrightarrow \quad \left(\Pi^T - I_n\right) \pi^* = 0$$

We know this object. Given a transition matrix Π, this equation tells us that its associated stationary distribution π^* will be given by the eigenvector of Π^T associated to its unitary eigenvalue. Of course, we may have more than just one unitary eigenvalue and therefore more than one stationary distribution.

Example 1 *A markov chain characterized by the transition matrix*

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ .2 & .7 & .1 \\ 0 & 0 & 1 \end{bmatrix}$$

has two unitary eigenvalues corresponding to the distributions $\pi_a^{*T} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$ and $\pi_b^{*T} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$. The states $i = 1$ and $i = 3$ are called absorbing states because it is impossible to leave them. The vectors π_a^* and π_b^* describe two stationary distributions, and so does any linear combination of them.

Existence and uniqueness of the stationary distribution are useful properties. We can state some sufficient conditions for them.

Definition 3 *A Markov chain is called an ergodic chain if it is possible to go from every state to every other state (not necessarily in one move)*

Definition 4 *Let Π be a transition matrix satisfying property (2). We call the corresponding Markov chain regular if $\left(\Pi^T\right)^n_{i,j} > 0 \ \forall (i,j)$ for some $n \geq 1$*

Notice that the notation $\left(\Pi^T\right)^n_{i,j}$ refers to the (i,j)th element of the matrix $\left(\Pi^T\right)^n$. Therefore, a Markov chain is called a *regular* chain if some power of the transition matrix has only positive elements.

According to these two definitions, any regular chain will be ergodic but it is not necessarily true that any ergodic chain will be regular.

Example 2 *A markov chain characterized by the transition matrix*

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

is ergodic because it is possible to move from any state into the other one. However, it is not regular: in odd powers of Π it is not possible to stay in the given state and in even powers of Π it is not possible to leave the given state.

It is quite self-evident, but we can also state that an ergodic matrix does not have any absorbing state.

Theorem 1 *Let Π be a transition matrix. If property (2) is satisfied there exists at least one unitary eigenvalue of Π and therefore there exists at least one stationary vector π^* associated to Π that will satisfy property (1)*

Theorem 2 *Let Π be a transition matrix satisfying property (2). If Π satisfies the definition of a regular chain then,*

- *Π has a unique stationary distribution π^**
- *Given any $n \times 1$ vector π satisfying property (1), we have*

$$\lim_{n\to\infty} \left(\Pi^T\right)^n \pi = \pi^*$$

For instance, any transition matrix with no zero will have a unique stationary distribution. But still, we may have transition matrices with zeros such that there is a unique stationary distribution. This will occur whenever we can move from any point in the state space into any other point in n steps.

3 Discretizing a continuous process

Let's assume we have a random variable z_t that follows the following $AR(1)$ process:

$$z_t = \rho z_{t-1} + \epsilon_t \quad \text{with } 0 \leq \rho < 1 \tag{3}$$

with

$$\epsilon_t \sim N\left(0, \sigma_\epsilon^2\right) \tag{4}$$

where ρ and σ_ϵ^2 are either known or estimated from data.

We need a discrete version for this process. A natural mathematical object that will fit our need is a Markov chain. A Markov chain is characterized by two objects, a vector $Z \in \mathbb{R}^n$ that describes the n possible states and an $n \times n$ transition matrix Π that describes the probability of moving from one state to another in the following way: the element $\Pi_{i,j}$ is the probability of moving from state i to state j. We ask this matrix to satisfy property (2), that is to say, rows must add up to one. Therefore, we need to choose the values of Z and Π that *best* approximate the continuous $AR(1)$ process above. We proceed as follows:

1. The variance of the process z_t is given by:

$$\sigma_z^2 = \frac{\sigma_\epsilon^2}{1 - \rho^2}$$

   Then, we choose the upper and lower bounds of our discrete process to be $m\sigma_z$ and $-m\sigma_z$, where m is an arbitrary number that will determine the amplitude of the state space. Therefore, $Z_1 = -m\sigma_z$ and $Z_n = m\sigma_z$. If we want an equally spaced grid we proceed by defining:

$$dz = \frac{Z_n - Z_1}{n - 1}$$

   and any element i of our vector Z will be:

$$Z_i = Z_1 + dz\,(i - 1)$$

2. We approximate the element Π_{ij} of the transition matrix by the probability of moving from Z_i into the interval $\left(Z_j - \frac{dz}{2}, Z_j + \frac{dz}{2}\right]$ for $j \neq 1, n$.[1] This is a relatively straightforward computation:

$$\begin{aligned} \Pi_{ij} &\simeq Pr\left(Z_j - \frac{dz}{2} < z_{t+1} \leq Z_j + \frac{dz}{2} | z_t = Z_i\right) \\ &= Pr\left(z_{t+1} \leq Z_j + \frac{dz}{2} | z_t = Z_i\right) - Pr\left(z_{t+1} \leq Z_j - \frac{dz}{2} | z_t = Z_i\right) \end{aligned}$$

[1] Obviously, for $j = 1$ we consider the interval $\left(-\infty, Z_1 + \frac{dz}{2}\right]$ and for $j = n$ the interval $\left(Z_n - \frac{dz}{2}, \infty\right)$

Then, using equation (3) we can rewrite

$$\Pi_{ij} \simeq Pr\left(\epsilon_{t+1} \leq Z_j + \frac{dz}{2} - \rho Z_i\right) - Pr\left(\epsilon_{t+1} \leq Z_j - \frac{dz}{2} - \rho Z_i\right)$$

Using our distributional assumption (4) we can rewrite:

$$\Pi_{ij} \simeq \phi\left[\frac{Z_j + \frac{dz}{2} - \rho Z_i}{\sigma_\epsilon}\right] - \phi\left[\frac{Z_j - \frac{dz}{2} - \rho Z_i}{\sigma_\epsilon}\right]$$

where $\phi(\cdot)$ is the *cdf* for the standard normal distribution.

4 Random walks

Now, let's take the process in the previous section for the case in which $\rho = 1$. We have a random variable z_t that follows the following process:

$$z_t = z_{t-1} + \epsilon_t \tag{5}$$

with

$$z_0 \sim N\left(0, \sigma_{z_0}^2\right) \quad \text{and} \quad \epsilon_t \sim N\left(0, \sigma_\epsilon^2\right) \tag{6}$$

where $\sigma_{z_0}^2$ and σ_ϵ^2 are either known or estimated from data. Therefore, the process for z_t is normally distributed with zero mean and variance given by

$$Var\left(z_t\right) = \sigma_{z_0}^2 + t\,\sigma_\epsilon^2 \tag{7}$$

Note that this process is not stationary. The variance of z_t increases with t and tends to infinity as t tends to infinity. Therefore, as a process for labor earnings it will only be well defined when applied to an OG model where people live for a finite number of periods T.

We need a discrete version for this process. We will use a series of T Markov chains characterized by T vectors $Z_t \in \mathbb{R}^n$ and T transitions matrices $\Pi_t \in \mathbb{R}^n \times \mathbb{R}^n$. Note that we hold fixed the dimension n of the series of chains. We proceed as follows:

1. Let's call $\sigma_{z_t}^2$ the variance of the stochastic process described in equation (7). Then, we choose the upper and lower bounds of our vector Z_t to be $m\sigma_{z_t}$ and $-m\sigma_{z_t}$, where m is an arbitrary number that will determine the amplitude of the state space. Therefore, $Z_{t,1} = -m\sigma_{z_t}$ and $Z_{t,n} = m\sigma_{z_t}$. If we want an equally spaced grid we proceed by defining:

$$dz_t = \frac{Z_{t,n} - Z_{t,1}}{n-1}$$

and any element i of our vector Z_t will be:

$$Z_{t,i} = Z_{t,1} + dz_t \left(i - 1\right)$$

2. We approximate the element $\Pi_{t,ij}$ of the transition matrix by the probability of moving from $Z_{t,i}$ into the interval $\left(Z_{t+1,j} - \frac{dz_{t+1}}{2}, Z_{t+1,j} + \frac{dz_{t+1}}{2}\right]$ for $j \neq 1, n$.[2] This is a relatively straightforward computation:

$$\begin{aligned}\Pi_{t,ij} &\simeq Pr\left(Z_{t+1,j} - \frac{dz_{t+1}}{2} < z_{t+1} \leq Z_{t+1,j} + \frac{dz_{t+1}}{2} | z_t = Z_{t,i}\right) \\ &= Pr\left(z_{t+1} \leq Z_{t+1,j} + \frac{dz_{t+1}}{2} | z_t = Z_{t,i}\right) - Pr\left(z_{t+1} \leq Z_{t+1,j} - \frac{dz_{t+1}}{2} | z_t = Z_{t,i}\right)\end{aligned}$$

Then, using equation (5) we can rewrite

$$\Pi_{t,ij} \simeq Pr\left(\epsilon_{t+1} \leq Z_{t+1,j} + \frac{dz_{t+1}}{2} - Z_{t,i}\right) - Pr\left(\epsilon_{t+1} \leq Z_{t+1,j} - \frac{dz_{t+1}}{2} - Z_{t,i}\right)$$

Using our distributional assumption in expression (6) we can rewrite:

$$\Pi_{t,ij} \simeq \phi\left[\frac{Z_{t+1,j} + \frac{dz_{t+1}}{2} - Z_{t,i}}{\sigma_\epsilon}\right] - \phi\left[\frac{Z_{t+1,j} - \frac{dz_{t+1}}{2} - Z_{t,i}}{\sigma_\epsilon}\right]$$

where $\phi\left(\cdot\right)$ is the *cdf* for the standard normal distribution.

Note that although the discretization of the process is very similar to the stationary case ($\rho < 1$), its implementation is somehow more cumbersome: we need as many vectors Z_t and transition matrices Π_t as periods in our model.

[2] Obviously, for $j = 1$ we consider the interval $\left(-\infty, Z_{t+1,1} + \frac{dz_{t+1}}{2}\right]$ and for $j = n$ the interval $\left(Z_{t+1,n} - \frac{dz_{t+1}}{2}, \infty\right)$