316-406 ADVANCED MACROECONOMIC TECHNIQUES NOTE 4b

Chris Edmond ⟨cpedmond@unimelb.edu.au⟩

Again, we will adopt the following notational conventions: **random variables** will be denoted by capital letters, like X_t and Z_t, **realizations** of random variables will be denoted by corresponding little letters, say x_t and z_t, a **stochastic process** will be a sequence of random variables, say $\{X_t\}$ and $\{Z_t\}$, and a **sample path** will be a sequence of realizations, say $\{x_t\}$ and $\{z_t\}$.

Markov chains

Roughly speaking, a stochastic process $\{X_t\}$ has the **Markov property** if the probability distributions

$$\Pr(X_{t+1} \leq x | x_t, x_{t-1}, \cdots x_{t-k}) = \Pr(X_{t+1} \leq x | x_t)$$

for any $k \geq 2$. A Markov chain is a stochastic process with this property and which takes values in a **finite set**. A Markov chain (x, P, π) is characterized by a a triple of three objects: a **state space** identified with an n-vector x, an n-by-n **transition matrix** P, and an **initial distribution**, an n-vector π_0.

Let

$$x = (x_1, \cdots, x_n)$$

Then the transition matrix $P = [p_{ij}]$ has elements with the interpretation

$$p_{ij} = \Pr(X_{t+1} = x_j | X_t = x_i)$$

So, fix a row i. Then the elements in each of the j columns give the conditional probabilities of transiting from state x_i to state x_j. In order for these to be well-defined probabilities, we require

$$1 \geq p_{ij} \geq 0, \qquad i, j = 1, 2, ...n$$

and for each i,

$$\sum_{j=1}^{n} p_{ij} = 1, \qquad i = 1, 2, ...n$$

For example, if $n = 2$ and

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{10} & \frac{9}{10} \end{pmatrix}$$

then the conditional probability of moving from state $i = 1$ to $i = 2$ is $p_{12} = \frac{1}{2}$ while the conditional probability of moving from state $i = 2$ to $i = 1$ is only $p_{21} = \frac{1}{10}$. Notice that the diagonal elements of P give a notion of the persistence of the chain, because these elements give the probability of staying in a given state. As another example, consider

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

This has the property that if you are in state $i = 1$, you can move to state $i = 2$ with probability $p_{12} = \frac{1}{2}$ but if you are in $i = 2$, you are stuck there. In this case, $i = 2$ is said to be an **absorbing state**.

Similarly, the initial distribution $\pi_0 = [\pi_{0,i}]$ has elements with the interpretation

$$\pi_{0,i} = \Pr(X_0 = x_i)$$

And in order for these to be well-defined probabilities, we require

$$1 \geq \pi_{0,i} \geq 0, \qquad i = 1, 2, ...n$$

and

$$\sum_{i=1}^{n} \pi_{0,i} = 1$$

In applications, we often set the initial distribution to have $\pi_i = 1$ for some i and zero everywhere else so that the chain is started in state i with probability 1.

A. Higher order transitions

The main convenience of the Markov chain model is the ease with which they can be manipulated using ordinary linear algebra. For example,

$$\begin{aligned} \Pr(X_{t+2} = x_j | X_t = x_i) &= \sum_{k=1}^{n} \Pr(X_{t+2} = x_j | X_{t+1} = x_k) \Pr(X_{t+1} = x_k | X_t = x_i) \\ &= \sum_{k=1}^{n} p_{kj} p_{ik} \\ &= (P^2)_{ij} \end{aligned}$$

where $(P^2)_{ij}$ is the ij element of the matrix $P^2 = PP$. In general, $(P^2)_{ij} \neq p_{ij}^2$ where the latter is the square of the ij element of P. More generally,

$$\Pr(X_{t+s} = x_j | X_t = x_i) = (P^s)_{ij}$$

The Markov chain gives a **law of motion for probability distributions** over a finite set of possible state values. The sequence of unconditional probability distributions is $\{\pi_t\}_{t=0}^{\infty}$ where each π_t is an n-vector. Let $\pi_t = [\pi_{t,i}]$ be a vector whose elements have the interpretation

$$\pi_{t,i} = \Pr(X_t = x_i)$$

Then the sequence of probability distributions are computed using

$$\begin{aligned} \pi_1' &= \pi_0' P \\ \pi_2' &= \pi_1' P = \pi_0' P^2 \\ &\vdots \\ \pi_t' &= \pi_{t-1}' P = \pi_0' P^t \end{aligned}$$

where a prime ($'$) denotes the transpose. In short, probability distributions π_t' evolve according to the linear homogeneous system of difference equations

$$\pi_{t+1} = P' \pi_t$$

We already know a lot about solving systems of such difference equations.

The conditional and unconditional moments of the Markov chain can be calculated using similar reasoning. For example,

$$\begin{aligned} \mathsf{E}\{X_1\} &= \pi_0' x \\ \mathsf{E}\{X_2\} &= \pi_0' P x \end{aligned}$$

and

$$\mathsf{E}\{X_t\} = \pi_0' P^t x$$

while conditional expectations are

$$\mathsf{E}\{X_{t+k}|X_t = x\} = P^k x$$

Similarly, let g be a n-vector that represents a function with typical element $g(x)$ on the state x. Then

$$\mathsf{E}\{g(X_{t+k})|X_t = x\} = P^k g$$

B. Stationary distributions

A steady-state or **stationary probability distribution** is a vector $\bar{\pi}$ such that

$$\bar{\pi}' = \bar{\pi}' P$$

or equivalently

$$(P' - I)\bar{\pi} = 0$$

Recall that a scalar λ is an **eigenvalue** of P' if and only if $P' - \lambda I$ is a singular matrix. A square matrix P' is **singular** if and only if its **determinant** is zero. Hence λ is an eigenvalue of P' if and only if

$$\det(P' - \lambda I) = 0$$

An equivalent way of saying that a matrix A is **non-singular** is to say that the only solution of the equation $Ay = 0$ is $y = 0$. Equivalently, P' is singular if and only if $Ay = 0$ has solutions other than $y = 0$. Hence if λ is an eigenvalue of P' such that $A = P' - \lambda I$ is singular, there must be solutions other than $y = 0$ to the equation $Ax = (P' - \lambda I)y = 0$. Equivalently, if λ is an eigenvalue of P', there must be at least one vector $y \neq 0$, called an **eigenvector**, such that

$$P'y = \lambda y, \qquad y \neq 0$$

In the context of our Markov chain model, a stationary probability distribution is an eigenvector $y = \bar{\pi}$ associated with a unit eigenvalue $\lambda = 1$ of P'. The requirement that $\sum_{i=1}^{n} \bar{\pi}_i = 1$ is a normalization of the eigenvector.

Because $1 \geq p_{ij} \geq 0$ and $\sum_{i=1}^{n} p_{ij} = 1$, any transition matrix P must have **at least one** unit eigenvalue $\lambda = 1$, but there may be more than one such eigenvalue. Hence there may be more than one stationary distribution. For example, if

$$P = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{5} & \frac{1}{2} & \frac{3}{10} \\ 0 & 0 & 1 \end{pmatrix}$$

then the matrix P has two unit eigenvalues (one corresponding to the 1 in the first row, the other corresponding to the 1 in the third row) and the associated Markov chain has two stationary distributions, namely $\bar{\pi}_1 = [1, 0, 0]$ and $\bar{\pi}_2 = [0, 0, 1]$.

Moreover, even if a Markov chain has a unique stationary distribution, there is no guarantee that the chain will asymptotically converge to that unique stationary distribution for all initial conditions. A sufficient condition for the existence of a unique asymptotically stable stationary distribution is that $1 > p_{ij} > 0$ for each i, j so that the chain has no absorbing states.

Chris Edmond

22 August 2004