Notes on Discrete-Time Markov Processes

François Gourio*

These notes are based on the references quoted in the last section.

1. Motivation

Many interesting models are recursive if exogenous variables (shocks) follow Markov processes. In this case, the endogenous (state and control) variables themselves follow Markov processes. To derive analytical or numerical results about these variables we need to analyze Markov processes. Concretely, the savings problem

$$\begin{aligned} V(a,y) &= \max_{c\geq 0, a'\geq 0} \left\{ U(c) + \beta \mathbb{E}_{y'/y} V(a',y') \right\}, \\ a' &= Ra + y - c, \\ & \quad y \text{ Markov}, \end{aligned}$$

leads to a policy function $a' = g(a,y)$. This implies that the vector $X_t = (a_t, y_t)$ follows a Markov process. We'll present tools that allow, for instance, to make predictions regarding the average level of assets $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} a_t$.

2. Markov Chains: basic facts and definitions

Definition 1. *A stochastic process (in discrete time) is a sequence of random variables:* $(X_t)_{t\geq 0}$ *where for all* $t = 0, 1, 2, ...,$ X_t *is a random variable.*

Example 1. *Any linear ARMA process; or a nonlinear process, e.g.* $X_t = \exp(\varepsilon_t)\, \varepsilon_{t-1}$ *with* ε_t *iid* $N(0, \sigma^2)$.

Definition 2. *A stochastic process satisfies the Markov property if for all* $t \geq 0$,

$$\Pr(X_{t+1} \mid X_t) = \Pr(X_{t+1} \mid X_t, ..., X_0).$$

*Boston University, Dept of Economics. Email: fgourio@bu.edu.

Example 2. *A random walk* $X_{t+1} = X_t + \varepsilon_{t+1}$, ε_t *iid* $N(0, \sigma^2)$;*or* $X_{t+1} = \cos(t) + \rho X_t + \varepsilon_{t+1}$;*or* $X_{t+1} = F(X_t, \theta_{t+1})$ *with* θ_{t+1} *an iid variable.*

Exercise 1. *Are martingales markov processes? Is the reverse true? (Recall a martingale is a stochastic process such that* $X_t = \mathbb{E}_t X_{t+1}$, *all* $t \geq 0$.*)*

Definition 3. *A stochastic process is a (time-homogeneous) Markov chain if (1) it satisfies the Markov property, (2)* $\Pr(X_{t+1} \mid X_t)$ *does not depend on time* t *and (3) for all* t, X_t *takes values in a finite set* S. *Hence, a Markov chain is given by (a) a finite state space* S *where* X_t *takes it values; (b) a* $S \times S$ *transition matrix* Q, *with* $\forall x, y \in S : Q(x,y) = \Pr(X_{t+1} = y \mid X_t = x)$; *and (c) a distribution* μ_0 *over* S *from which the initial value* X_0 *is drawn in* S. *(This distribution could be degenerate, meaning we start for sure in some given state.)*

We will denote the state space as the set of integers between 1 and $S : \{1, 2, ..., S\}$.

The matrix Q *must satisfy* $Q(x,y) \geq 0$ *for all* $x, y \in S$, *and* $\sum_{y \in S} Q(x,y) = 1$ *for all* $x \in S$. *A matrix satisfying these conditions is called a stochastic matrix.*

μ_0 *must be a distribution on* $\{1, 2, ..., S\}$, *i.e.* $\mu_0(s) \geq 0$ *for all* $s = 1...S$ *and* $\sum_{s \in S} \mu_0(s) = 1$. *The set of distributions on* $\{1, ..., S\}$ *is called the simplex of* $\mathbb{R}^S$ *and is denoted* $\Delta^S = \left\{ p \in \mathbb{R}^S_+, \ \sum_{s=1}^S p_s = 1 \right\}$.

Remark 1. *Be careful since there is also a notation* $Q(j \mid i)$ *to mean* $\Pr(X_{t+1} = j \mid X_t = i)$.

Remark 2. *Below I discuss briefly the case of Markov process in discrete time with a general state space: this means we would not require the condition (3). As an example of a discrete, countable state space:* $X_{t+1} = X_t + \varepsilon_{t+1}$ *with* $\varepsilon_{t+1} = +1$ *with probability* $1/2$ *and* $\varepsilon_{t+1} = -1$ *with probability* $1/2$.

Example 3. *(detailed).* $S = \{1, 2\}$, $\mu_0 = [1, 0]'$, $Q = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$. *This matrix* Q *means that when you are in state* 1, *with probability* $3/4$ *you stay in state 1, and with probability* $1/4$ *you go to state* 2. *When you are in state 2, with probability* $1/4$ *you go to state* 1 *and with probability* $3/4$ *you stay in state* 2. *Note that the* $i-th$ *row gives the probability of going tomorrow to each state, conditional on being in state* i *today. Hence, the sum of each row must be* 1. *Inversely, the column* j *gives the conditional probabilities of going to* j, *conditional on being in states* $1, 2.., S$. *The distribution* μ_0 *indicates we start for sure in state* 1.

Example 4. *$S = \{1,2\}$, $\mu_0 = [1,0]'$, $Q = \begin{bmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \end{bmatrix}$ for some $\alpha \in [0,1]$. This is an iid process: not only is $\Pr(X_{t+1} \mid X_t, ..., X_0) = \Pr(X_{t+1} \mid X_t)$ independent of $X_{t-1}, ..., X_0$, it is also independent of X_t : $\Pr(X_{t+1} \mid X_t) = \Pr(X_{t+1})$.*

Example 5. *$S = \{1,2\}$, $\mu_0 = [1,0]'$, $Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This is a deterministic sequence. If t is odd, $X_t = 2$, and if t is even, $X_t = 1$.*

Example 6. *$S = \{1,2,3\}$, $\mu_0 = [0,0,1]'$, $Q = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 3/4 & 0 \\ 0 & 0.1 & 0.9 \end{bmatrix}$. Notice how after some time the system will leave state 3 and never return to it.*

Example 7. *$S = \{1,2,3\}$, $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 2/3 & 1/3 \end{bmatrix}$. Notice how depending on the initial condition, the system may either get stuck in state 1, or oscillate stochastically b/w state 2 and state 3.*

Example 8. *(Mehra and Prescott, 1985 JME). They model US consumption growth by the following two-state process: $S = \{1,2\}$, $\mu_0 = [1,0]'$, $Q = \begin{bmatrix} \frac{1}{2}+p & \frac{1}{2}-p \\ \frac{1}{2}-p & \frac{1}{2}+p \end{bmatrix}$. State 1 is a "boom" with consumption growth high and state 2 is a "recession" with consumption growth low: if $x(s)$ is consumption growth, we choose values for $x(1)$ and $x(2)$ appropriately: $x(1) = 1+\mu+\varepsilon$, $x(2) = 1+\mu-\varepsilon$. (these are annual real gross rates of growth). If $p > 0$ there is some persistence in consumption growth. MP set $\mu = 0.019$, $\varepsilon = 0.035$ and $p = -0.28$. (Example to be continued).*

Remark 3. *As always,*

$$\Pr(X_t, X_{t-1}, ..., X_0) = \Pr(X_t \mid X_{t-1}, ..., X_0) \Pr(X_{t-1} \mid X_{t-2}, ..., X_0) ... \Pr(X_1 \mid X_0)\mu(X_0).$$

Now given the Markov assumption, we have the simplified formula:

$$\begin{aligned} \Pr(X_t, X_{t-1}, ..., X_0) &= \Pr(X_t \mid X_{t-1}) \Pr(X_{t-1} \mid X_{t-2}) ... \Pr(X_1 \mid X_0)\mu(X_0), \\ &= Q(X_{t-1}, X_t) Q(X_{t-2}, X_{t-1}) ... Q(X_0, X_1)\mu(X_0). \end{aligned}$$

(This makes computing likelihoods easy.)

Remark 4. *Graphical representation of a Markov chain.*

Proposition 1. *Let X_t be a Markov Chain (S, μ_0, Q). Then for any $k \geq 1$,*

$$\Pr\left(X_{t+k} = j \mid X_t = i\right) = (Q^k)_{i,j},$$

where $(Q^k)_{i,j}$ is the i, j element of the matrix Q to the power k.

Proof: we prove this by induction. For $k = 1$ the formula is the definition of Q. Assume the formula is true for some $k \geq 1$:

$$\Pr\left(X_{t+k} = j \mid X_t = i\right) = \left(Q^k\right)_{i,j}.$$

Then we use the total probability formula to write:

$$\Pr\left(X_{t+k+1} = j \mid X_t = i\right) = \sum_{y \in S} \Pr\left(X_{t+k+1} = j \mid X_{t+k} = y, X_t = i\right) \Pr\left(X_{t+k} = y \mid X_t = i\right).$$

Next we use the Markov property, the definition of Q, and the induction assumption to rewrite this as

$$\Pr\left(X_{t+k+1} = j \mid X_t = i\right) = \sum_{y \in S} Q(y, j) \left(Q^k\right)_{i,y}.$$

But we recognize this as the formula for the product of the matrix Q by the matrix Q^k, hence

$$\Pr\left(X_{t+k+1} = j \mid X_t = i\right) = \left(Q^{k+1}\right)_{i,j},$$

and the result is proven for $k + 1$.$\square$

We now turn to the "accounting" of probabilities. Suppose the distribution at time 0 over the states is drawn according to a probability μ_0, then what is the distribution over the states next period? The following result answers this question.

Proposition 2. *Let for all $k = 1 ... S$, $\pi_{t,k} = \Pr(X_t = k)$ be the unconditional probability of being in state k at time t, and let π_t be the (column) vector $(\pi_{t,1}, ..., \pi_{t,S})'$. Then:*

$$\forall t \geq 0, \pi_{t+1} = Q' \pi_t.$$

Proof:

$$\begin{aligned}
\forall s &= 1...S: \\
\pi_{t+1,s} &= \Pr\left(X_{t+1}=s\right) \\
&= \sum_{y\in S} \Pr\left(X_{t+1}=s \mid X_t=y\right)\Pr\left(X_t=y\right) \\
&= \sum_{y\in S} Q(y,s)\pi_{t,y} \\
&= \sum_{y\in S} \left(Q'\right)_{s,y}\pi_{t,y},
\end{aligned}$$

which proves our point since this is the $s-$*coordinate of the multiplication of* Q' *and* π_t.$\square$

An immediate corollary is that we can obtain the $k-$period ahead unconditional probabilities $\pi_{t+k} = \left(Q'\right)^k \pi_t$. In particular if we start at time 0 with a state drawn from μ_0, after t periods the distribution is $\pi_t = Q'^t\mu_0$. Clearly if the sequence of matrixes $\left(Q'^t\right)_{t\geq 0}$ converges to some matrix R, then $\pi_t \to R\mu_0$. In this case, we see that the process will tend in the long run to spend $\left(R\mu_0\right)(s)$ time in each state s. These long-run distributions are interesting because they reflect the "average" behavior of the system. We can try this numerically by computing $Q'^t\mu_0$ for large t, and you can try this by hand on the various examples we used above. As the examples show, it need not be the case that Q'^t converges to a matrix R (see also SLP 11.1 for the analysis of various examples). But in any case, if π_t converges, its limit must satisfy $\pi = Q'\pi$. Hence the

Definition 4. *π, a probability distribution on $\{1,...,S\}$ is invariant if $\pi = Q'\pi$.*

Remark 5. *The set of invariant probability distributions is a compact, convex subset of Δ^S. In particular, if there are two invariant distributions, there is an infinity of them.*

Remark 6. *Note the definition is equivalent to saying that π is an eigenvector associated to the eigenvalue* 1 *of the matrix Q'.*

In the next section, we prove that any Markov chain has an invariant distribution; but there may be several invariant distributions (take the example $Q = I$, the identity matrix); and finally the Markov process may not converge to any of these distributions. We will give sufficient conditions under which we have unicity, and sufficient conditions under which we have convergence.

Exercise 2. *Find the invariant distribution(s) of the Markov chains examples we gave above.*

Exercise 3. *Compute by hand the conditional mean and the conditional standard deviation of the examples above: what is* $E(X_{t+1} \mid X_t)$ *and* $Var(X_{t+1} \mid X_t)$*? (Obviously you will need to compute this for each state* X_t*.)*

Exercise 4. *Compute by hand the unconditional mean, standard deviation, and autocorrelation of the process used by Mehra and Prescott. Express the mean, standard deviation and autocorrelation as a function of* μ, ε, p*. To do this you will need the invariant distribution of the process* $\{X_t\}$*.*

Remark 7. *Simulating a markov chain is easy given a random number generator. Given the state* $x(t) \in S$*, you need to use the conditional distribution* $[Q(x(t),1),...,Q(x(t),S)]$ *to draw* $x(t+1)$*. Draw a number* $s(t)$ *between* 0 *and* $1,$ *and if* $s(t) < Q(x(t),1)$*, then* $x(t+1) = 1$*; if* $Q(x(t),1) \leq s(t) < Q(x(t),1) + Q(x(t),2)$*, then* $x(t+1) = 2$*; and so on. Once you know* $x(t+1)$ *you can repeat the process with the new distribution* $[Q(x(t+1),1),...,Q(x(t+1),S)]$*, and so on to get a time series of the desired length.*

3. Long-run behavior of Markov Chains

3.1. General discussion

I am merely rephrasing here the introduction of chapter 11 which is a nice summary (also see chapter 2, and the introduction of chapter 8 to see how everything hangs together).

Recall the analysis of the neoclassical growth model without uncertainty.[1] Once you have optimized, you obtain a law of motion for capital along the optimal path: $k_{t+1} = g(k_t)$.

Three questions arise naturally:

1. Existence of a steady-state: is there a number k^* such that $k^* = g(k^*)$?

2. Unicity: is there only one such k^*?

[1] To wit, $\max_{c_t,k_{t+1}} \sum_{t\geq 0} \beta^t U(c_t)$ s.t. $k_{t+1} = (1-\delta)k_t + F(k_t) - c_t$, $k_0 > 0$ given.

3. Stability: do we converge to this steady-state, starting from any initial condition k_0? i.e. do we have $\lim_{t\to\infty} k_t = k^*$, for any k_0? This is global stability. We can weaken this by requiring only that this property hold for any k_0 that is near enough k^*. This is local stability.

Given a Markov process with transition matrix Q, we have stochastic sample path of X_t, hence there will typically not be convergence to a constant, but we can ask whether the X_t's will settle on average around some area. This leads us to ask the questions:

1. Is there an invariant distribution μ such that $T^*\mu = \mu$?

2. Is it unique?

3. Do we have convergence to this stationary distribution from any initial distribution, i.e. do we have that $\lim_{t\to\infty} (T^*)^t \mu_0 = \mu^*$ for any μ_0.

3.2. Results

We consider a Markov chain (S, μ_0, Q). The aim is to prove the theorems 11.1 and 11.4 in SLP, but the presentation is slightly different.

Definition 5. *We say that state $j \in S$ is accessible from state $i \in S$, and we note $i \hookrightarrow j$, if there is a positive probability, starting from i, to get to j, i.e.: $\exists N \geq 0, \pi_{i,j}^{(N)} > 0$, where we note $\pi_{i,j}^{(N)} = \Pr(X_N = j \mid X_0 = i)$, i.e. the (i,j) element of Q^N.*

Remark 8. *The opposite of $i \hookrightarrow j$ is $\forall n \geq 0, \pi_{i,j}^{(n)} = 0$, i.e. starting from i, one will never go to j.*

Proposition 3. *If $i \hookrightarrow j$ and $j \hookrightarrow k$ then $i \hookrightarrow k$.*

Proof: $\pi_{i,k}^{(n+m)} \geq \pi_{i,j}^{(n)} \pi_{j,k}^{(m)} > 0$.

Definition 6. *We say that state i is recurrent if for all j that is accessible from i, i is accessible from j, i.e.: $i \hookrightarrow j \Rightarrow j \hookrightarrow i$.*

Definition 7. *We say that state i is transient if it is not recurrent, i.e. there is a state j such that $i \hookrightarrow j$ but not$(j \hookrightarrow i)$. That is, there is a positive probability of going to j and never returning to i.*

Proposition 4. *If i is recurrent and $i \hookrightarrow j$, then j is recurrent.*

Proof: Let k be such that $j \hookrightarrow k$. Since $i \hookrightarrow j$, we have $i \hookrightarrow k$. Since i is recurrent, we have $k \hookrightarrow i$. Since $i \hookrightarrow j$, we have $k \hookrightarrow j$. Hence j is recurrent.

Definition 8. *We say that i and j communicate with each other if $i \hookrightarrow j$ and $j \hookrightarrow i$. From the results just proven, we note that "to communicate" is an equivalence relation, i.e (i) i communicates with i, (i) if i communicates with j then j communicates with i, and (iii) if i communicates with j and j communicates with k then i communicates with k.*

Definition 9. *An ergodic set is a subset E of S such that (i) $\Pr(X_n \in E \mid X_0 \in E) = 1$ for all $n \geq 0$, and (2) no strict subset of E has this property.*

Once you enter an ergodic set you never leave it. An ergodic set that has only one element is called an absorbing state: once you reach it, you remain stuck forever.

One way to explain the following theorem is to note that the Markov chain will after some time leave any transient state, since every time you leave it, there is some probability of not returning. Hence not all states can be transient, otherwise there would be no place to go after a while. The proof formalizes this idea.

Proposition 5. *Any Markov chain has at least one recurrent state.*

Proof: by contradiction, assume all states are transient. First, notice that $\pi_{i,i} < 1$ for all $i = 1...S$ otherwise the state i would be recurrent; hence, each state leads to at least one other state (i.e. some other state is accessible from each state). Next, start from state 1 : we know that 1 is transient, so there is some state, say 2, such that for some $n \geq 0$ $\pi^{(n)}_{1,2} > 0$ but for all $n \geq 0$, $\pi^{(n)}_{2,1} = 0$ (there is a probability of going there and never returning). Now, state 2 is transient too, so there is some state x such that $\pi^{(n)}_{2,x} > 0$ for some $n \geq 0$ and $\pi^{(n)}_{x,2} = 0$ for all $n \geq 0$. Note x cannot be 1 since $\pi^{(n)}_{2,1} = 0$. Call "3" the state x then. Note that 1 cannot be accessible from 3 either since $3 \hookrightarrow 1$ and $1 \hookrightarrow 2$ would imply $3 \hookrightarrow 2$, which is false. Hence by induction we have that for all $s \in S$, for all $n \geq 0$ and for all $i < s$, $\pi^{(n)}_{s,i} = 0$. But this leads to a contradiction for $s = S, n = 1$ since $\pi_{S,i} = 0$ for all $i = 1...S-1$ and $\pi_{S,S} < 1$ as we noted as the outset: we don't have $\sum_{j \in S} \pi_{S,j} = 1$. Contradiction.$\square$

Proposition 6. *One can partition the state space S into a set T of transient states and $M \geq 1$ sets $E_1, ..., E_M$, which all contain only recurrent states, with each E_i an ergodic set.*

Proof: put all the transient states in T; then among the recurrent states put two in the same set iff they communicate with each other (by definition of a recurrent state, if one state is accessible from another, they communicate with each other). By construction, the sets are ergodic: the probability of going to T is 0 (since the states in the E_i are recurrent, there would be a probability of return, contradicting that T has only transient states); moreover the probability of going to E_j for $j \neq i$ is 0 too since we created the sets by putting states that communicate with each other together. Hence, $\Pr(X_{t+1} \in E_i \mid X_t \in E_i) = 1$ for all i. And clearly no proper subset of the E_i will verify this, since it will leave out some state with is accessible from the other states. (What we did is partition the set of recurrent states into the equivalence classes of the relation "to communicate with each other".)$\square$

Proposition 7. *$\frac{1}{N}\sum_{k=0}^{N-1} Q^k$ converges as $N \to \infty$ to a matrix R which is stochastic.*

Proof: Define $T^N = \frac{1}{N}\sum_{k=0}^{N-1} Q^k$. Since all the matrixes Q^k have their elements taking values in $[0,1]$, we see that T^N is also a matrix which elements take values in $[0,1]$. Hence the sequence $\left\{T^N\right\}$ lives in a compact space (since bounded and closed finite-dimensional). We can thus extract a convergent subsequence $T^{\phi(N)}$. Call $R = \lim T^{\phi(N)}$. Clearly R is also a stochastic matrix (take the limit in the definition of a stochastic matrix). Now we want to prove that T^N actually converges to R. First notice that $QT^N = T^N Q = \frac{1}{N}\sum_{k=1}^{N} Q^k = T^N + \frac{-I+Q^N}{N}$ hence by taking the limit for $N = \phi(n)$, $n \to \infty$, we have $QR = RQ = R$. The same argument with Q^i instead of Q yields that for any given i, $Q^i R = RQ^i = R$. Now assume that T^N doesn't converge to R; since T^N lives in a compact, we can extract a convergent subsequence $T^{\psi(n)}$ converging to a different limit $\overline{R}$. The same argument implies that $\overline{R}Q^i = Q^i\overline{R} = \overline{R}$ for all i. But this implies in particular that $\overline{R}T^{\phi(n)} = T^{\phi(n)}\overline{R} = \overline{R}$, and letting $n \to \infty$ leads to $\overline{R}R = R\overline{R} = \overline{R}$. Applying the same result to $RT^{\psi(n)} = T^{\psi(n)}R = R$, we get $R\overline{R} = \overline{R}R = R$. Hence $R = \overline{R}$. Hence T^N must converge to R.$\square$

This theorem implies that the average probability distribution over the states converges. The average time spent in each state after N iterations is indeed $T^N \mu_0$.

Proposition 8. *There always exists an invariant distribution. Moreover, any row of R is an invariant distribution, and any invariant distribution is a convex combination of the rows of R.*

Proof: from the previous theorem, we know that $\frac{1}{N}\sum_{k=0}^{N-1} Q^k \to R$, *a stochastic and* $RQ = R$. *Writing the equality of these matrixes as the equality of the row vectors, we have* $r_{s-} = [r_{s1}, ..., r_{sS}] = r_{s-} \times Q$ *so each row* r_{s-} *is an invariant distribution. Moreover, an invariant distribution* π *satisfies*

$$\begin{aligned} \forall n &\geq 1 : \pi(s) = \sum_{i \in S} Q^n(i,s)\pi(i) \\ \pi(s) &= \sum_{i \in S} \left(\frac{1}{N} \sum_{k=0}^{N-1} Q^k \right)(i,s)\pi(i) \to \sum_{i \in S} R(i,s)\pi(i), \end{aligned}$$

which shows that π *is a convex combination of the rows of* R *(with the coefficients the* π *themselves!).*

Another proof of the first point: Use Brouwer's fixed point theorem: a continuous mapping from a compact, convex subset of $\mathbb{R}^S$ *into itself has a fixed point. The mapping is* $\pi \to Q'\pi$.□

This gives a way of finding numerically the invariant distribution: compute T^N for large N and take the rows. (As we will see below, in the usual case one can even do simpler and compute Q^N only.) Another way is to ask directly the computer to find the eigenvalues and eigenvectors of Q'. (In this case of course you need to normalize the eigenvectors to make them probability distributions.)

However, we are still left with the case where there are multiple invariant distributions. The following theorem gives a necessary and sufficient condition for the existence of a unique ergodic set, which is also a sufficient condition for the uniqueness of an invariant distribution.

Proposition 9. *Q has a unique ergodic set iff there is a state j such that* $\forall i \in S, \exists n \geq 1, \pi_{i,j}^{(n)} > 0$. *Moreover in this case Q has an unique invariant distribution* π.

Proof: If such a state j exists, it is recurrent; since it is accessible from all states, there is at most one ergodic set. Since we know there is one from the result above, there is only one. Since each invariant distribution is associated with an ergodic set, there is a unique invariant distribution. Reverse clear (or see SLP).□

Definition 10. *The Markov chain* (S, μ_0, Q) *has cyclically moving subsets if* $\exists C_1, C_2 \subset S$ *with* $\Pr(X_{n+1} \in C_2 \mid X_n \in C_1) = 1$, $\Pr(X_{n+1} \in C_1 \mid X_n \in C_2) = 1$ *and* $C_1 \cap C_2 = \emptyset$.

Proposition 10. *Let for all* $n \geq 1$*,all* $j \in S$*,* $\varepsilon_j^{(n)} = \min_i \left(\pi_{ij}^{(n)}\right)$ *and let* $\varepsilon^{(n)} = \sum_{j\in S} \varepsilon_j^{(n)}$*. Then the Markov chain* (S, μ_0, Q) *has a unique ergodic set with no cyclically moving subset iff* $\exists N \geq 1, \varepsilon^{(N)} > 0$*. In this case, for any* $\mu_0 \in \Delta^S$*,* $Q'^N \mu_0 \to \pi$*, the unique invariant distribution (and convergence occurs at a geometric rate).*

The condition is clearly the same as: $\exists j \in S, \exists N \geq 0, \forall i \in S, \pi_{i,j}^{(N)} > 0$*. Note that the key element here is that the* N *is the same for all* i, j*. This is the difference with the previous proposition where the probability of going to* j *was nonzero for some* N*, which could depend on* i*. By considering the example 5 (a periodic chain), you can understand why this assumption will rule out the case of cyclically moving subset. In that case, the condition if 'satisfied' for different* N*'s, depending on the* i*, and it is not possible to find a same* N *that works for all the* i*'s. Hence the theorem doesn't work, one only has the existence of a unique invariant distribution.*

Remark 9. *Note that a (weak) sufficient condition for this theorem is that for some* $N \geq 0$*,* $\pi_{i,j}^{(N)} > 0$ *for all* $i, j \in S$*.*

In this case, which is the most common in applications, computing Q^N for large N is enough to find the invariant distribution. One will find that the limiting matrix has identical rows, which are all the invariant distribution.

Some more definitions and results

I state without proof the following result which is intuitive.

Proposition 11. i *is recurrent iff* $\Pr\left[X_n = i \text{ for an infinity of } n \mid X_0 = i\right] = 1$*, and* i *is transient iff* $\Pr\left[X_n = i \text{ for an infinity of } n \mid X_0 = i\right] = 0$.

Definition 11. *A Markov chain is irreducible if it has only one ergodic set (all states communicate with each other).*

(To do: add more results. add more examples.)

4. More general state space

In some cases the state space is not a finite set. It may be countable but not finite, or it may be continuous. The results we proved in sections 1 and 2 then do not hold in general, they require stronger assumptions. In this section I follow

SLP (section 8.1). The more general notation encompasses our finite state case as a special case.

Let $(S, \Im, \mathbb{P})$ be a probability space.[2]

Definition 12. *A transition function is a function Q from $S \times B(\Im) \to \mathbb{R}$ such that*

(1) for any $s \in S$, $Q(s, .)$ is a probability measure on $(S, \Im)$.

(2) for any $A \in \Im$, $Q(., A)$ is a measurable function from $(S, \Im)$ into $\mathbb{R}$.

$Q(s, A)$ should be interpreted as $\Pr(X_{t+1} \in A \mid X_t = s)$.

Example 9. *In the case of a Markov chain with transition matrix P, $Q(i, \{j\}) = P(i, j)$. Hence we define $Q(s, A)$ for any $A \subset \{1, ...S\}$ as $Q(s, A) = \sum_{i \in A} P(s, i)$.*

Example 10. *Suppose that $X_t \in \mathbb{R}$ and we define $F(x \mid s) = \Pr(X_{t+1} \leq x \mid X_t = s)$. Then $Q(s, A) = \int_A dF(x \mid s)$.*

Example 11. *A normal AR(1) process: $X_{t+1} = \rho X_t + \varepsilon_{t+1}$ with ε_t iid $N(0, \sigma^2)$. We have*

$$\begin{aligned} Q(s, A) &= \Pr(X_{t+1} \in A \mid X_t = s) \\ &= \Pr(\rho s + \varepsilon_{t+1} \in A) \end{aligned}$$

Hence we can define for any b :

$$Q(s,]-\infty, b]) = \Pr(\rho s + \varepsilon \leq b) = \Pr(\varepsilon \leq b - \rho s) = \Phi_{0,1}\left(\frac{b - \rho s}{\sigma}\right),$$

where $\Phi_{0,1}$ is the standard normal cdf. [Having defined Q for any interval $]-\infty, b]$ is enough: it allows us to define $Q(s, A)$ for any $A \in B(\Im)$, with the formula $Q(s, A) = \int_{-\infty}^{+\infty} 1_{u \in B} g(s, u) du$, where $g(s, u) = dQ(s,]-\infty, b])/db\rfloor_{b=u} = \frac{1}{\sigma}\phi_{0,1}\left(\frac{u-\rho s}{\sigma}\right)$.]

Let $\Lambda(S, \Im)$ the set of probability measures over $(S, \Im)$ and $M(\Im)$ the set of measurable functions from $(S, \Im)$ into $\mathbb{R}$. We now define two operators associated with the Markov chain Q.

[2] Recall this means that S is a set, $\Im$ is a sigma-algebra over S, and $\mathbb{P}$ is a probability measure over $(S, \Im)$. A sigma-algebra is a set of parts of S satisfying (i) $\emptyset \in \Im$, (ii) $A \in \Im \Rightarrow A^c \in \Im$, and (iii) if A_n is a countable collection of sets in $\Im$, then $\cup_{n \in \mathbb{N}} A_n$ is in $\Im$. A probability measure $\mathbb{P}$ on $(S, \Im)$ is a mapping from $\Im$ into $[0, 1]$ such that: (i) $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(S) = 1$, (ii) if the A_n are countable and disjoint, then $\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.

Definition 13. *The Markov operator T is a mapping from the set of measurable functions $(S, \Im) \to \mathbb{R}$ into itself defined as:*

$$\forall f \in M(S, \Im), \forall s \in S : (Tf)(s) = \int f(s')Q(s, ds').$$

Interpretation: $(Tf)(s) = \mathbb{E}\left(f(s_{t+1}) \mid s_t = s\right).$

Remark 10. *T is linear: $T(\alpha f + \beta g) = \alpha Tf + \beta Tg$ for any $f, g : (S, \Im) \to \mathbb{R}$ and $\alpha, \beta \in \mathbb{R}$.*

Remark 11. *In the case of Markov chains, $(Tf)(s) = \sum_{s' \in S} Q(s, s')f(s')$ which we can write in vector terms as $Tf = Qf$, where f is the column vector $\{f(s)\}_{s=1...S}$ and Tf the vector $\{(Tf)(s)\}_{s=1...S}$.*

SLP show in theorem 8.1 and its corollary that T maps positive measurable functions into positive measurable functions, and bounded measurable functions into bounded measurable functions: $T : M^+(S, \Im) \to M^+(S, \Im)$, and $T : B(S, \Im) \to B(S, \Im)$.

Example 12. *Let's write down T for the example of an AR(1). We have*

$$\begin{aligned} (Tf)(s) &= \int f(s')Q(s, ds') \\ &= \int_{-\infty}^{+\infty} f(s') \frac{1}{\sigma} \phi_{0,1} \left(\frac{s' - \rho s}{\sigma} \right) ds'. \end{aligned}$$

Definition 14. *The adjoint of T is T^*, a mapping from the set of probability measures over $(S, \Im)$ into itself, defined as:*

$$\forall \lambda \in \Lambda(S, \Im), \forall A \in \Im : (T^*\lambda)(A) = \int Q(s, A)\lambda(ds).$$

Interpretation: $T^*\lambda$ is next period's probability distribution over the states if λ is today's probability distribution over the states: $(T^*\lambda)(A) = \Pr(s_{t+1} \in A \mid s_t \rightsquigarrow \lambda)$.

Remark 12. *T^* is satisfies $T^*(\alpha\lambda + (1 - \alpha)\mu) = \alpha T^*\lambda + (1 - \alpha)T^*\mu$ for any $\alpha \in [0, 1], \lambda, \mu \in \Lambda(S, \Im)$.*

Remark 13. *In the case of Markov chains,* $T^*\mu = \mu Q$.

Example 13. *In the case of an AR(1),*

$$\begin{aligned} (T^*\lambda)(A) &= \int Q(s,A)\lambda(ds) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} 1_{u\in A}\frac{1}{\sigma}\phi_{0,1}\left(\frac{u-\rho s}{\sigma}\right) du\,\lambda(ds) \end{aligned}$$

Or, for some $b \in \mathbb{R}$ *and* $A =]-\infty, b]$:

$$\begin{aligned} (T^*\lambda)(]-\infty,b]) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} 1_{u\leq b}\frac{1}{\sigma}\phi_{0,1}\left(\frac{u-\rho s}{\sigma}\right) du\,\lambda(ds) \\ &= \int_{-\infty}^{+\infty} \Phi_{0,1}\left(\frac{b-\rho s}{\sigma}\right)\lambda(ds). \end{aligned}$$

Finally if λ *has a density i.e.* $\lambda(ds) = g(s)ds$, *then*

$$(T^*\lambda)(]-\infty,b]) = \int_{-\infty}^{+\infty} \Phi_{0,1}\left(\frac{b-\rho s}{\sigma}\right) g(s)ds$$

Definition 15. *We can define the* $n-$*period probability ahead recursively. For any* $s \in S, A \in \Im$, *let* $Q^1(s,A) = Q(s,A)$, *and* $\forall n \geq 1, Q^{n+1}(s,A) = \int Q^n(u,A)Q(s,du)$.

Proposition 12. *For all* $n \geq 1$, Q^n *is a transition function, and the operators of the iterate are given as the iterates of the operator of the one-step ahead transition function:* $T^{(n)} = T^n, T^{*(n)} = T^{*n}$.

Proof: see SLP.

Remark 14. *Hence, given an initial distribution* $\lambda_0, \lambda_n = T^{*n}\lambda_0$ *is the distribution over the states after* n *periods. Moreover* $T^n f$ *is the expected value of* f *n-periods in advance:* $(T^n f)(s) = \mathbb{E}(f(s_{t+n}) \mid s_t = s)$.

Example 14. *Consider again the case of an AR(1). Let* G_{t+1} *be the c.d.f. and* g_{t+1} *the p.d.f. of* X_{t+1}. *Then we have the recursion,* $\forall x \in \mathbb{R}$:

$$\begin{aligned} G_{t+1}(x) &= \Pr(X_{t+1} \leq x \mid X_t \rightsquigarrow G_t) \\ G_{t+1}(x) &= \int_{-\infty}^{+\infty} \Phi_{0,1}\left(\frac{x-\rho s}{\sigma}\right) G_t(ds) \\ G_{t+1}(x) &= \int_{-\infty}^{+\infty} \Phi_{0,1}\left(\frac{x-\rho s}{\sigma}\right) g_t(s)ds \end{aligned}$$

Hence:

$$\forall x \in \mathbb{R} : g_{t+1}(x) = \frac{1}{\sigma} \int_{-\infty}^{+\infty} \phi_{0,1}\left(\frac{x-\rho s}{\sigma}\right) g_t(s)ds,$$

and a stationary distribution must satisfy the equation $T^*\lambda = \lambda$, with we can write in p.d.f. terms as

$$\forall x \in \mathbb{R} : g^*(x) = \frac{1}{\sigma} \int_{-\infty}^{+\infty} \phi_{0,1}\left(\frac{x-\rho s}{\sigma}\right) g^*(s)ds.$$

This is a functional equation in g^*. In our case, we can guess and verify that the stationary distribution of X will be normal (μ, Σ^2). We see that μ, Σ^2 have to satisfy: $\mathbb{E}X_{t+1} = \rho\mathbb{E}X_t + \mathbb{E}\varepsilon_{t+1}, \mathbb{V}X_{t+1} = \rho^2\mathbb{V}X_t + \sigma^2 \Rightarrow \mu = 0, \Sigma^2 = \sigma^2/(1-\rho^2)$. To check this guess, we need to see that $g^*(x) = \phi_{0,1}\left(\frac{x-\mu}{\Sigma}\right)/\Sigma$ satisfies the equation above, i.e. we need to check that

$$\phi_{0,1}\left(\frac{x-\mu}{\Sigma}\right) = \frac{1}{\sigma} \int_{-\infty}^{+\infty} \phi_{0,1}\left(\frac{x-\rho s}{\sigma}\right) \phi_{0,1}\left(\frac{s-\mu}{\Sigma}\right) ds,$$

which is possible to do by hand. In this special case, there is also another way, which is simpler: given the properties on the sum and product by a scalar of normal distributions, we see that if X_t is normal (μ, Σ), then X_{t+1} is normal (μ, Σ). This suffices to prove that this is one invariant distribution.

More generally, the definition of an invariant distribution is still $T^*\mu = \mu$, i.e.:

$$\forall A \in \Im : \int Q(s, A)\mu(ds) = \mu(A),$$

which we can first rewrite by requiring this to hold only for the subsets $A =]-\infty, b]$, denoting $F(b \mid s) = Q(s,]-\infty, b]) = \Pr(X_{t+1} \leq b \mid X_t = s)$:

$$\forall b \in \mathbb{R} : \int F(b \mid s)\mu(ds) = \mu(]-\infty, b]),$$

and in the special case when (i) F has a continuous density i.e. $F(b \mid s) = f(b \mid s)ds$, and (ii) μ has a continuous density too, i.e. $\mu(ds) = g(s)ds$, we can rewrite this equality as

$$\forall b \in \mathbb{R} : \int f(b \mid s)g(s)ds = g(b).$$

Example 15. *There is a simple formula for* $\mathbb{E}\left(\sum_{t\geq 0}\beta^t f(x_t)\mid x_0\right)=\sum_{t\geq 0}\beta^t\left(T^t f\right)(x_0)$ *where* T^t *denotes the* $t-th$ *iterate of* T *in the case of Markov chains:* $\sum_{t\geq 0}\beta^t\left(T^t f\right)(x_0)=\left(\left(\sum_{t\geq 0}\beta^t Q^t\right)f\right)(x_0)=e(x_0)(I-\beta Q)^{-1}f$ *where* $e(x_0)$ *is the* $1\times S$ *vector with* 1 *in position* x_0 *and* 0 *elsewhere.*

Remark 15. *We can obtain the expression for* $\mathbb{E}\left(\sum_{t\geq 0}\beta^t f(x_t)\mid x_0\right)$ *by a slightly different road. The distribution of* x_t *given* x_0 *is* $\left(T^{*t}\right)(x_0)=e(x_0)Q^t$. *Next,* $\mathbb{E}\left(f(x_t)\mid x_0\right)=e(x_0)Q^t f$ *where* f *is the* $S\times 1$ *vector of values* $f(s)$. *Finally,*

$$\mathbb{E}\left(\sum_{t\geq 0}\beta^t f(x_t)\mid x_0\right)=\sum_{t\geq 0}\beta^t e(x_0)Q^t f=e(x_0)(I-\beta Q)^{-1}f.$$

What we do in this example is compute of the future distribution using T^* instead of computing the expected value of f using T. This is an instance of the following theorem.

Proposition 13. $\forall f\in M^+(S,\Im),\forall\lambda\in\Lambda(S,\Im):$

$$\int (Tf)(s)\lambda(ds)=\int f(s)(T^*\lambda)(ds).$$

Proof: see SLP (Theorem 8.3).

This simply says that if today's state is drawn according to λ, the expected value of f tomorrow can be found by first finding the probability distribution tomorrow if today's state is drawn according to λ, and then integrating f with respect to this distribution. In the case of Markov chains we already know this since the equality can be written: for any $f\in\mathbb{R}^S$ (a $S\times 1$ vector) for any $\lambda\in\Delta^S$ (a $S\times 1$ vector):

$$\lambda'(Qf)=(\lambda' Q)f,$$

which is obviously true.

Remark 16. *At the risk of beating a dead horse, there is a third way to compute the discounted sum of the earlier example. Let* $V(x_0)=\mathbb{E}\left(\sum_{t\geq 0}\beta^t f(x_t)\mid x_0\right)$, *then*

$$\begin{aligned} V(x_0) &= f(x_0)+\beta\mathbb{E}\left(V(x_1)\mid x_0\right)\\ V(x_0) &= f(x_0)+\beta\sum_{s\in S}Q(x_0,s)V(s)\end{aligned}$$

which we can write in vector terms as

$$V=f+\beta QV\Rightarrow V=(I-\beta Q)^{-1}f.$$

Example 16. *(Campbell and Cochrane, JPE 1999). Assume log consumption is a random walk:* $\Delta \log C_t = \mu + \varepsilon_t$ *with* ε_t *iid* $N(0, \sigma^2)$. *CC specify a utility function* $\frac{1}{1-\sigma}\mathbb{E}\sum_{t\geq 0}\beta^t (C_t - X_t)^{1-\sigma}$ *where* X_t *depends implicitely on past consumption and today's consumption as follows. Let* $S_t = (C_t - X_t)/C_t$ *and let* $s_t = \log S_t$ *satisfy for some chosen parameter values* $\phi, \overline{s}$:

$$\begin{aligned}
s_{t+1} &= (1-\phi)\overline{s} + \phi s_t + \lambda(s_t)\left(\Delta \log C_{t+1} - \mu\right), \\
\lambda(s) &= \frac{1}{e^{\overline{s}}}\sqrt{1 - 2\left(s - \overline{s}\right)} - 1, \ \ s \leq s_{\max} \\
&= 0, \ \ s \geq s_{\max} \\
s_{\max} &= \overline{s} + \frac{1}{2}\left(1 - e^{2\overline{s}}\right).
\end{aligned}$$

Write the transition function associated with this Markov process. Find the stationary distribution numerically. What is interesting about this example is that it is nonlinear, and the nonlinearities are important.

We finish with two important definitions:

Definition 16. *Q has the Feller property if its Markov operator T maps $C(S)$, the set of continuous function of S into $\mathbb{R}$, into itself, i.e. if the conditional expectations are continuous in today's state:* $(Tf)(s) = \int f(s')Q(s, ds')$ *is continuous in s for any f continuous.*

Small variations in the state today move only a little the conditional expectation tomorrow of any continuous function; intuitively this means that the probability distribution over states tomorrow moves only a little, and since f is continuous, the conditional expectation moves only a little too.

Definition 17. *Q is monotone if its associated Markov operator T maps nondecreasing functions into nondecreasing functions.*

This means that the higher the state today, the higher the conditional expectation of tomorrow's f, for any f. This is equivalent to saying that for any $s_2 \geq s_1$, $Q(s_2, .)$ first-order stochastically dominates $Q(s_1, .)$.

[Recall that a c.d.f. F first-order stochastically dominates G iff any of the following equivalent conditions are satisfied: (i) $\forall x \in \mathbb{R}, F(x) \leq G(x)$, (ii) for any h weakly increasing, $\int h(s)dF(s) \geq \int h(s)dG(s)$, (iii) X drawn from F equals Y drawn from G plus a random variable $Z \geq 0$; of course a consequence that F FOSD G is that the mean of F is higher than the mean of G.]

5. Convergence results for Markov Processes

Chapters 11 and 12 in SLP study the convergence properties of the more general Markov processes in detail. Here I only quote two important results.

Proposition 14. *If $S \subset \mathbb{R}^l$ is compact and Q has the Feller property, then there exists an invariant distribution.*

Proposition 15. *Assume that $S = [a, b]$, Q is monotone, has the Feller property, and satisfies the following mixing condition:*

$$\exists c \in S, \exists \varepsilon > 0, \exists N \geq 1, P^N(a, [c, b]) \geq \varepsilon \text{ and } P^N(b, [a, c]) \geq \varepsilon.$$

Then there is a unique invariant distribution μ^* and for any initial distribution μ_0, the sequence $\mu_n = (T^*)^n \mu_0$ converges weakly (i.e. in distribution) to μ^*.

6. References

- Stokey, Lucas and Prescott, Recursive Methods in Economic Dynamics, Harvard UP. Sections 11.1 and 8.1 especially.
- Ljungqvist and Sargent, Recursive Macroeconomic Theory, MIT Press, chapter 1 has a simpler section on Markov chains.
- A math book that is quite readable is "Basic Stochastic Processes" (Brzezniak and Zastawniak, Springer Undergraduate texts). There are many other good probability books.