Chapter 12

Thursday, July 28, 2011

12.1 Markov Processes

Today's lecture will focus on the central role that Linear Algebra plays in understanding *probabilistic* or *stochastic* processes, where a system evolves according to certain probabilities. In particular, the kinds of processes that we will focus on are those where the state of a system only depends on the state of the system in the previous time step. These systems, called *Markov processes*, have far-reaching applications in dynamics, economics, and statistical mechanics.

12.1.1 Motivating Markov Processes - An Example from Sociology

Sociologists are interested in group dynamics and, in particular, how individuals flow into and out of various subpopulations of society. We will consider a simplistic scenario where there are exactly two subpopulations and individuals can be in exactly one of these. The setting is the Greater Los Angeles area; this sprawl consists of two types of inhabitants: those that live in *urban* settings (e.g., downtown LA) and those in *suburban* settings (e.g., Pasadena). A sociologist has computed that there is much movement every census period (10 years) between the two regions of Los Angeles. Specifically, she has computed that every ten years, 40% of those living in urban settings move to the suburbs (and thus 60% of those in urban settings remain urban). Furthermore, she has found that every ten years, 30% of the suburban dwellers move to an urban setting (and thus 70% of suburbanites remain suburban).

Using these statistics, we may compute what the suburban population S and the urban population U is every ten years. So, assume that at the 0-th census reading, we have subpopulations $S_0 = 1000$ and $U_0 = 3000$. The above statistics indicate that to calculate the urban population U_1 at the first census reading, we must add the number of individuals staying urban (which will be $.6 \cdot 3000$) and the number of people switching from suburban to urban (which will be $.3 \cdot 1000$). Thus, at the first census reading, we obtain the subpopulation

$$U_1 = .6 \cdot 3000 + .3 \cdot 1000 = 1800 + 300 = 2100.$$

In a similar fashion, we may calculate the suburban population at S_1 at the first census by adding those that remain suburban ($.7 \cdot 1000$) to those that convert from urban life to suburban life ($.4 \cdot 3000$). Thus, the total suburban population at the first census reading will be

$$S_1 = .4 \cdot 3000 + .7 \cdot 1000 = 1200 + 700 = 1900.$$

If we abstract this calculation a bit, we begin to see where concepts from Linear Algebra are beginning to crop up. In particular, if we do not specify the initial subpopulations S_0 and U_0 and leave them as these variables, we obtain the following system of equations in solving for S_1 and U_1:

$$\begin{aligned} U_1 &= .6U_0 + .3S_0 \\ S_1 &= .4U_0 + .7S_0 \end{aligned} .$$

The linearity of systems begs us to evoke matrix multiplication to write this system as

$$\begin{bmatrix} U_1 \\ S_1 \end{bmatrix} = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix} \begin{bmatrix} U_0 \\ S_0 \end{bmatrix}.$$

Of course, we are also interested what happens at the second census. Since we are assuming that the rate at which individuals move from S to U is constant, we see that we may generalize the above matrix equation. In fact, if we want to know what the subpopulations $[U_{k+1}, S_{k+1}]$ are at the $(k+1)$-st census, we only need to know the subpopulations $[U_k, S_k]$ at the k-th census. These two sets of subpopulations are related by the matrix equation

$$\begin{bmatrix} U_{k+1} \\ S_{k+1} \end{bmatrix} = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix} \begin{bmatrix} U_k \\ S_k \end{bmatrix}.$$

We also note that $[U_k, S_k]$ can be computed by knowing $[U_{k-1}, S_{k-1}]$. If we continue along this fashion, we see that the only piece of information we really need is what the initial subpopulations $[U_0, S_0]$ was. In fact, we may find $[U_k, S_k]$ by multiplying our initial population vector by our transition matrix k times. Thus, we have the following matrix equation for the subpopulations at census k:

$$\begin{bmatrix} U_k \\ S_k \end{bmatrix} = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix}^k \begin{bmatrix} U_0 \\ S_0 \end{bmatrix}.$$

12.1.2 Analyzing our Markov Process

The simplicity of this situation obviously means that this result does not mirror reality. In particular, our system assumes that no one is born or dies. Furthermore, it assumes that an individual will be in exactly one of the two states S or U when in fact more complicated living situations exists. One of the other unrealistic assumptions is that the rate at which subpopulations flow from one state to another remains the same at every census period. In fact, U.S. historical analyses indicate that these rate flow coefficients have changed dramatically in the past several decades.

This example does, though, give us an excellent starting example of a Markov process. A ***discrete Markov process*** is a system with n states evolving over a discrete time parameter t, where the system at time t depends only on the system at time $t-1$. We see that our sociological example is indeed a Markov process with $n = 2$ states (one is either urban or suburban) because the subpopulations at the k-th census reading only depend on the subpopulations at the $(k-1)$-st census reading.

To generalize our example a bit, if a system has n states, then we can represent the system at time k by a vector $\mathbf{S_k} = [S_{k1}, S_{k2}, \ldots, S_{kn}]$, where S_{kn} indicates the number of objects in state n at time k. To describe how these states change, we give the probability p_{ij} that an object in state i will transition to state j in the next time step. So, if we want to know how many objects are expected in state j at time $k+1$, we need to compute

$$S_{k+1,j} = p_{1j}S_{k1} + p_{2j}S_{k2} + \cdots + p_{nj}S_{kn}.$$

Doing this for all j, we obtain the following matrix equation

$$\begin{bmatrix} S_{k+1,1} \\ S_{k+1,2} \\ \vdots \\ S_{k+1,n} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & p_{31} & \cdots & p_{n1} \\ p_{12} & p_{22} & p_{32} & \cdots & p_{n2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{1n} & p_{2n} & p_{3n} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} S_{k1} \\ S_{k2} \\ \vdots \\ S_{kn} \end{bmatrix}.$$

This matrix, which we will call A is called a ***stochastic matrix*** or a ***transition matrix***. Notice that its ij entry is the probability p_{ji}, so the indexing of the probability and the row-column indexing are opposite. As this describes how the n states flow into and out of each other, this matrix will be a square $(n \times n)$-matrix. Of crucial important is that the individual entries are all probabilities. Thus, for every i and j, we have that $0 \leq p_{ij} \leq 1$. Furthermore, since these p_{ij} is the probability that an object will flow from state i to state j, we must have that

$$p_{i1} + p_{i2} + \cdots + p_{in} = 1,$$

and thus the entries of every column must sum to 1.

12.1.3 Steady State Vectors

When we are analyzing a system evolving over a time, we would like to know which configurations of our system do not change when we apply our transition matrix. These configurations, called ***steady states*** are analytically crucial to understanding the long-term dynamical behavior. So, assume that we have some stochastic $(n \times n)$-matrix A that gives us the transition probabilities between our n states. If an n-vector $\mathbf{x} \in \mathbb{R}^n$ is a steady state vector, then A will have no effect on $\mathbf{x}$ and thus we have the familiar equation

$$A\mathbf{x} = \mathbf{x}.$$

This equation is simply the eigenvalue-eigenvector equation with eigenvalue $\lambda = 1$. Of course, not all matrices have an eigenvalue 1, but perhaps since our stochastic matrix is so constrained (e.g., all terms between 0 and 1 and all columns summing to 1), there may always be a steady state vector. A famous theorem called the Perron-Frobenius theorem guarantees us that this is always the case:

Theorem - The Perron-Frobenius Theorem. Let A be a square stochastic matrix. Then A always has 1 as an eigenvalue and every other eigenvalue λ has $|\lambda| < 1$.

The existence of 1 as an eigenvalue ensures that there is always some nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{x}$; this $\mathbf{x}$ is exactly the steady-state vector we were hoping for.

Returning to our sociological example, recall that our Markov system was defined by

$$\begin{bmatrix} U_k \\ S_k \end{bmatrix} = \begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix}^k \begin{bmatrix} U_0 \\ S_0 \end{bmatrix}.$$

The Perron-Frobenius theorem guarantees that our stochastic matrix A has eigenvalue 1. We verify this computationally by finding the characterstic polynomial to be $\lambda^2 - 1.3\lambda + .30$. Using the quadratic equation, we get that the eigenvalues are

$$\lambda = \frac{1.3 \pm \sqrt{1.3^2 - 4 \cdot 1 \cdot .3}}{2} = \frac{1.3 \pm \sqrt{.49}}{2} = \frac{1.3 \pm .7}{2} = 1, .3.$$

As promised, we have an eigenvalue of 1 and can thus compute our steady state vectors by finding the eigenspace corresponding to the eigenvalue $\lambda = 1$. To do this, we solve $(A - I)\mathbf{x} = \mathbf{0}$. and get the equation

$$(A - I)\mathbf{x} = \begin{bmatrix} -.4 & .3 \\ .4 & -.3 \end{bmatrix} \begin{bmatrix} U_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives us a system of redundant equations, each of which is equivalent to $-.4U_0 + .3S_0 = 0$. Solving for U_0, we get that $U_0 = .75S_0$ and thus all steady state vectors have the form

$$\begin{bmatrix} U_0 \\ S_0 \end{bmatrix} = \begin{bmatrix} .75S_0 \\ S_0 \end{bmatrix} = S_0 \begin{bmatrix} .75 \\ 1 \end{bmatrix}.$$

Thus, any initial population configuration where the initial urban subpopulation U_0 is $3/4$ of the initial suburban population S_0 will give a system that remains constant over time. To verify this, we consider the initial subpopulation configuration given by $[3000, 4000]$. Applying our stochastic matrix, we see that

$$\begin{bmatrix} .6 & .3 \\ .4 & .7 \end{bmatrix} \begin{bmatrix} 3000 \\ 4000 \end{bmatrix} = \begin{bmatrix} 3000 \\ 4000 \end{bmatrix}.$$

Of course, since $A\mathbf{x} = \mathbf{x}$ for this particular $\mathbf{x}$, $A^k\mathbf{x} = \mathbf{x}$ as well. Thus, once a steady vector, always a steady state vector.

12.1.4 Long-Term Dynamics of Markov Processes

Another important feature of these evolving (dynamical) systems that we wish to understand is what happens in the long-run. That is, what happens if we allow the time in our dynamical system to tend towards infinity. For this to happen, we must compute

$$\mathbf{x}_\infty = \lim_{k \to \infty} A^k \mathbf{x}.$$

For this to even be possible, we must of course understand how to compute A^k for large values of k. Unfortunately, it is not true in general that the entries of A^k are obtained by taking the k-th power of each entry. Notice, though, that if had a diagonal matrix D (i.e. a square matrix with 0's off the diagonal and arbitrary numbers in the diagonal), then it is in fact true that D^k is the diagonal matrix obtained by simply raising each entry to the k-th power.

This begs the question of how we may transform the matrix A (which is not necessarily diagonal) to a diagonal matrix D. The concept that we reach is that of *change of basis*. Essentially, we want to change our basis from the standard one to a basis of eigenvectors. We want this because A acts on these eigenvectors by simply stretching them by their corresponding eigenvalues. Thus, if we change the basis for A to one given by eigenvectors, it will act like a diagonal matrix D, where the diagonal elements are exactly the eigenvalues of A.

To change basis, we need to find some invertible matrix P that has as its columns the new basis that we want. Of course, since the columns will form a basis, we are guaranteed that P will be invertible. For our purposes, the columns will eigenvectors. Finding this P, we may then change basis to have A act like D; the relationship between A, D, and P is given by the following change of basis equation:

$$P^{-1}AP = D.$$

Of course, the upshot of having this relationship is that we can then solve for A and obtain $A = PDP^{-1}$. This helps us move towards our goal of finding a nice form for A^k because adjacent P and P^{-1} will cancel out:

$$A^k = (PDP^{-1})^k = PDP^{-1}PDP^{-1} \cdots PDP^{-1}.$$

Canceling out the middle terms, we are left with

$$A^k = PD^kP^{-1}.$$

Of course, since D is diagonal, D^k is very easy to compute: just raise each diagonal element to the k-th power!

Returning to our sociology example, recall that we have already computed the eigenvalues of A to be 1 and .3. We also computed that any eigenvector for the eigenvalue $\lambda = 1$ is a scalar multiple of $[.75, 1]$. Computing the eigenvectors for $\lambda = .3$, we see that its eigenspace consists of scalar multiples of $[1, -1]$. So, to form P, we must choose one eigenvector for each eigenvalue (and we are free to choose whichever one we want). To avoid fractions, let us use $[3, 4]$ as the eigenvector for $\lambda = 1$ and $[1, -1]$ as the eigenvector for the eigenvalue $\lambda = .3$ We then have

$$P = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}.$$

We notice that we must compute the inverse of P. There is a well-defined method to do this for any invertible square matrix (e.g., using minors and cofactors). We present here the simple case of invertible an invertible (2×2)-matrix A:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Thus, the inverse of our matrix P is given by

$$P^{-1} = \frac{1}{7} \begin{bmatrix} 4 & -3 \\ 1 & 1 \end{bmatrix}.$$

One may quickly verify that indeed $P^{-1}AP = D$ by seeing that

$$P^{-1}AP = \frac{1}{7} \begin{bmatrix} 2.1 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} .3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Notice that the eigenvalue .3 is in the first spot diagonal spot because its corresponding eigenvector $[1, -1]$ is in the first column; a similar situation hold for the eigenvalue $\lambda = 1$.

We computed earlier that $A^k = PD^kP^{-1}$. Since D is diagonal, D^k is easy to compute and we are left with the following:

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} .3^k & 0 \\ 0 & 1^k \end{bmatrix} \frac{1}{7} \begin{bmatrix} 4 & -3 \\ 1 & 1 \end{bmatrix} =$$

$$\frac{1}{7} \begin{bmatrix} (4)(.3^k) + 3 & (-3)(.3^k) + 3 \\ (-4)(.3^k) + 4 & (.3)(.3^k) + 4 \end{bmatrix}.$$

If we take a limit as $k \to \infty$ of A^k and use the fact that $\lim_{k\to\infty} .3^k = 0$, we get

$$\lim_{k\to\infty} A^k = \frac{1}{7} \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix}.$$

Thus, in the limit, we expect that our subpopulations will tend towards

$$\lim_{k\to\infty} \frac{1}{7} \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} U_0 \\ S_0 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 3(U_0 + S_0) \\ 4(U_0 + S_0) \end{bmatrix}.$$

So, if our population had started off with $U_0 = 1000$ and $S_0 = 3000$, then it will eventually reach a limiting state of

$$\frac{1}{7} \begin{bmatrix} 12000 \\ 16000 \end{bmatrix}.$$

One surprising aspect about this limiting state is that it is an eigenvector for the eigenvalue $\lambda = 1$.