Stationary Distributions

Friday, February 21, 2014
2:02 PM

We now turn to developing formulas for solutions of Markov chains with properties that aren't restricted to a pre-specified finite time horizon.

- long-run properties
- questions about various possibilities that may happen "eventually"

To prepare to address these questions, the concept of a stationary distribution will be useful.

A stationary distribution of a Markov chain only is well-defined for time-homogenous Markov chains, and for such Markov chains is defined to be a vector

$\acute{\pi} = (\pi_1, \pi_2, \ldots, \pi_M)$

which satisfies the following conditions:

1. $\acute{\pi} P = \acute{\pi}$ (a left eigenvector of the probability transition matrix with eigenvalue 1)
2. $\pi_i \geq 0$
3. $\sum_{i=1}^{M} \pi_i = 1$

The last two conditions are simply the standard conditions on a probability distribution. Why is the first condition interesting?

Suppose that we initialize the Markov chain with an initial probability distribution given by a stationary distribution of that Markov chain.

Then if we denote $\phi_j^{(n)} = Prob(X_n = j)$

And represent the probability distribution of the state of the Markov chain at the nth epoch via the vector:

$\acute{\phi}^{(n)} = \left(\phi_1^{(n)}, \phi_2^{(n)}, \ldots, \phi_M^{(n)}\right)$

We are choosing $\acute{\phi}^{(0)} = \acute{\pi}$.

From our finite-horizon formula from last lecture,

$\acute{\phi}^{(n)} = \acute{\phi}^{(0)} P^n$

So by choosing the initial distribution to be the stationary distribution, we have:

$$\acute{\phi}^{(n)} = \acute{\pi}P^n = \acute{\pi}PP^{n-1} = \acute{\pi}P^{n-1} = \ \ldots = \acute{\pi}$$

In other words, if the state of the Markov chain is distributed according to the stationary distribution at one moment of time (say the initial epoch $n=0$), then the state of the Markov chain at any later time will still be governed by that same stationary distribution. So in some sense a stationary distribution is "natural" for a Markov chain in that the Markov chain dynamics preserve the stationary distribution. It's the probabilistic analogue of fixed points in deterministic dynamical systems. But it does not mean that the Markov chain gets stuck in a given state, just that the probability distribution doesn't evolve.

So we see one practical use of a stationary distribution is that if we initialize a Markov chain with a stationary distribution then the resulting stochastic process will behave as a "statistically stationary process" meaning that its statistics do not vary with time. But stationary distributions have much wider utility as we will now discuss.

Key mathematical issues related to stationary distributions, which have implications for application.

- Under what conditions do we have existence of a stationary distribution?
- Under what conditions do we have uniqueness of the stationary distribution?
- Under what conditions does the stationary distribution act as a limit distribution, meaning that arbitrary initial distributions will eventually, under the Markov chain dynamics, approach the stationary distribution? (global attractor)

Existence:

- Any finite-state Markov chain has a stationary distribution.
- Proof is based on the following observations:

$$P\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$$

  - Therefore the probability transition matrix has a right eigenvector with eigenvalue 1.
  - Therefore the probability transition matrix has a left eigenvector with eigenvalue 1, which therefore satisfies the first property of the stationary distribution.
  - The fact that this left eigenvector can be chosen to satisfy the other two properties of the stationary distribution follows from the Perron-Frobenius Theorem (Appendix B of Karlin and Taylor), which concerns properties of eigenvectors and eigenvalues of matrices with all

nonnegative entries.

Uniqueness:

To discuss the conditions under which a Markov chain has a unique stationary distribution, we will have to introduce some concepts involving the topology of the Markov chain.

Two states $i, j \in S$ of a Markov chain are said to communicate provided that there exist nonnegative integers m,n such that:

$$(P^n)_{ij} > 0 \text{ and } (P^m)_{ji} > 0.$$

Nonzero prob to go from i to j in exactly n steps

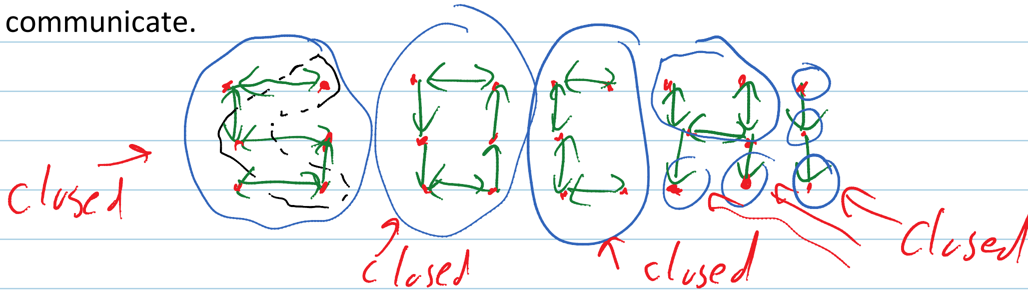
nonzero prob to go from j to i in exactly m steps

The intuitive interpretation of "communication" is simply that there is a way to go from state i to state j and vice versa via the Markov chain dynamics. The symbolic representation of communication is:

$i \leftrightarrow j$

The related notation $i \to j$ has obvious definition (possible to go from i to j, but not necessarily vice versa).

Communication is an equivalence relation so from generic algebra theorems, one can divide the state space into communication classes which have the property that all states within a communication class communicate, but that states in different communication classes don't communicate.



Another related definition: A communication class C is said to be closed if and only if there do not exist any states $i \in C$ and $k \notin C$ such that $i \to k$. In other words a communication class is closed if it has no outgoing connection with any other communication class.

A Markov chain is said to be irreducible if it has a single communication class (i.e., all states communicate with each other). Otherwise it is said to be reducible.

Note that even in a reducible Markov chain, any closed communication class can be meaningfully redefined as an irreducible Markov chain by simply taking the closed communication class as the new state space. This means that the statements we will make about irreducible Markov chains are also meaningful to apply to closed communication classes of reducible Markov chains.

Now we are prepared to state the uniqueness conditions along with an important formula.

Any irreducible finite-state Markov chain will have a unique stationary distribution which can be expressed by the following formula:

$$\pi_j = \left(\mathbb{E}|T_j(1)|X_0 = j|\right)^{-1}$$

where

$$T_j(1) = \min\{n: n > 0, X_n = j\}$$

is the first passage time into state j (or first return time) if, as here, one happens to start in state j. It is the random number of epochs required to visit state j after the initial epoch.

The meaning of the conditional expectation simply means to compute the expectation with a conditional probability distribution with the stated condition:

$$\mathbb{E}|T_j(1)|X_0 = j| = \sum_{n=0}^{\infty} n\, P(T_{j(1)} = n|X_0 = j)$$

Comments on the uniqueness result:

Why should there be the relationsihp between the stationary distribution and the first return time? Intuition: if the Markov chain has stationary distribution π then when it's in the stationary distribution, it will be in state j a fraction π_j of the time. Then what should be the average separation between visits to state j? $1/\pi_j$.

Example: $S = \{1, 2, 3\}$

$\pi_1 = .1, \quad \pi_2 = .7, \quad \pi_3 = .2$

1 2 2 2 3 2 2 2 2 1 2 2 3 2 2 2 3 2 2 2

This formula is usually useful in reverse, in the sense that (as we shall show in the next lecture), the stationary distribution is relatively easy to compute, and the formula tells you how it relates to the average first return time.

Is the irreducibility condition necessary for uniqueness? No; you can show fairly easily that uniqueness still applies to any Markov chain that has only one closed communication class. This is clearly necessary (else can have arbitrary choices of how to distribute the stationary distribution across the closed communication classes). Uniqueness still holds because one can show any stationary distribution must put zero weight on non-closed communication classes.

Finally, under what conditions does the stationary distribution serve as a limit distribution?

Need a further property of a Markov chain:

The period d(i) of a state i of a Markov chain is defined:

$d(i) = \gcd\{n > 0 : (P^n)_{ii} > 0\}$,

the greatest common divisor of the possible return times to state i.

Fortunately, period of a state is a class property in that it must be the same for all states in a communication class.

A state (and therefore class) is said to be aperiodic if it has period 1.

We'll see that aperiodicity together with irreducibility will imply an "ergodic property" that the stationary distribution is the limit distribution.