

Stationary Distributions

Monday, September 28, 2015
2:02 PM

No class on Thursday, October 1.

No office hours on Tuesday, September 29 and Thursday, October 1.

Homework 1 due Friday, October 2 at 5 PM

- strongly prefer printed copy in my mailbox in Amos Eaton 310
- acknowledge collaborations!

Last time we described how to compute arbitrary finite horizon statistics for a FSDT MC. But many questions of interest don't have a prescribed finite horizon:

- long run behavior
- questions about whether certain states are eventually achieved, and if so, how long does it take to happen

To begin to address these questions, we must develop some other characterizations and mathematical structures associated with Markov chains. We saw last time that we could compute finite horizon statistics for both time-homogenous and time-inhomogenous Markov chains. However, for indefinite horizon calculations, very few mathematical formulas are available for time-inhomogenous Markov chains, so we will **restrict consideration going forward to time-homogenous Markov chains**.

A **stationary distribution** of a time-homogenous FSDT Markov chain is defined to be a probability distribution which is preserved by the dynamics of the Markov chain. Specifically, it can be encoded as a (column) vector:

$$\vec{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_M \end{pmatrix}$$

with the following properties (where P is the probability transition matrix of the MC)

- $\pi^T P = \pi^T$
- $\pi_i \geq 0$

- $\sum_{i=1}^M \pi_i = 1$

The first property is the essential one:

Suppose that we define the probability distribution for the state of a FSDT MC after n epochs by the vector:

$$\vec{\phi}^{(n)} = \begin{pmatrix} p(\sum_n = 1) \\ p(\sum_n = 2) \\ \vdots \\ p(\sum_n = M) \end{pmatrix}$$

From last time, we know that the probability distribution for the state of FSDT MC updates as follows:

$$\vec{\phi}^{(n+1)T} = \vec{\phi}^{(n)T} P$$

(just refer to the calculation at end of last lecture
with $0 \rightarrow n, m \rightarrow n + 1$)

So, if $\vec{\phi}^{(n)} = \pi$, then $\vec{\phi}^{(n+1)} = \pi$ as well.

And in particular, if $\vec{\phi}^{(0)} = \pi$, then $\vec{\phi}^{(n)} = \pi$ for all $n \geq 0$.

So we see the analogy
stationary distribution : FSDT MCs

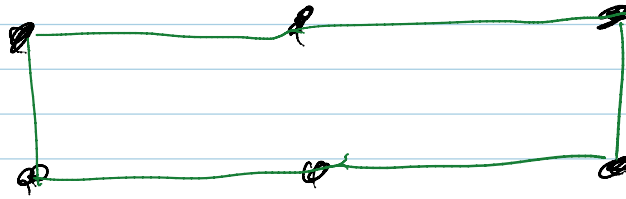
::

fixed points/equilibria : dynamical systems

In particular, the stationary distribution gives a probability distribution for the state of a FSDT MC which is determined by the dynamics, and so is somehow "natural."

Note that **what is fixed under the evolution of the stationary distribution are**

the probabilities, not the state variable itself which certainly changes.



Note that a FSDT MC which is initialized by a stationary distribution will give rise to a **statistically stationary** Markov process which means that the Markov process itself has statistics which are invariant under time translation

- to be distinguished by a general solution to a time-homogenous FSDT MC which has evolution rules that don't depend on time, but which can fail to be statistically stationary because the initial condition breaks time translation invariance.

Note also that an important modern example of the use of stationary distributions is in the PageRank algorithm, based on a "random surfer" model Markov chain for the network under consideration.

Key mathematical questions pertaining to stationary distributions:

- Under what conditions are we guaranteed that a stationary distribution for a FSDT MC **exists**?
- Under what conditions is the stationary distribution for a FSDT MC **uniquely determined**?
- Under what conditions does the stationary distribution serve as a **limit distribution**, meaning that the probability distribution for the Markov chain initialized arbitrarily will, after many epochs, approach the limit distribution?

Existence:

- Every FSDT MC is guaranteed to have a stationary distribution.
- The proof is based on the observation that, because all row sums of the probability transition matrix must $= 1$, then:

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

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

That means that P has a right eigenvector with eigenvalue 1.

Then linear algebra tells us P also has a left eigenvector with eigenvalue 1.

- equivalently observe that $I - P$ is singular

But then how do we guarantee the other two properties of stationary distribution?

- The **Perron-Frobenius Theorem** (Karlin and Taylor, Appendix B), which describes the properties of eigenvalues/eigenvectors of **matrices with nonnegative entries**, can be applied to guarantee that the left eigenvector can be chosen to have all nonnegative entries.
- Once we have a left eigenvector of eigenvalue 1 with nonnegative entries, we can just divide it by the sum of its entries so the normalization condition is satisfied.

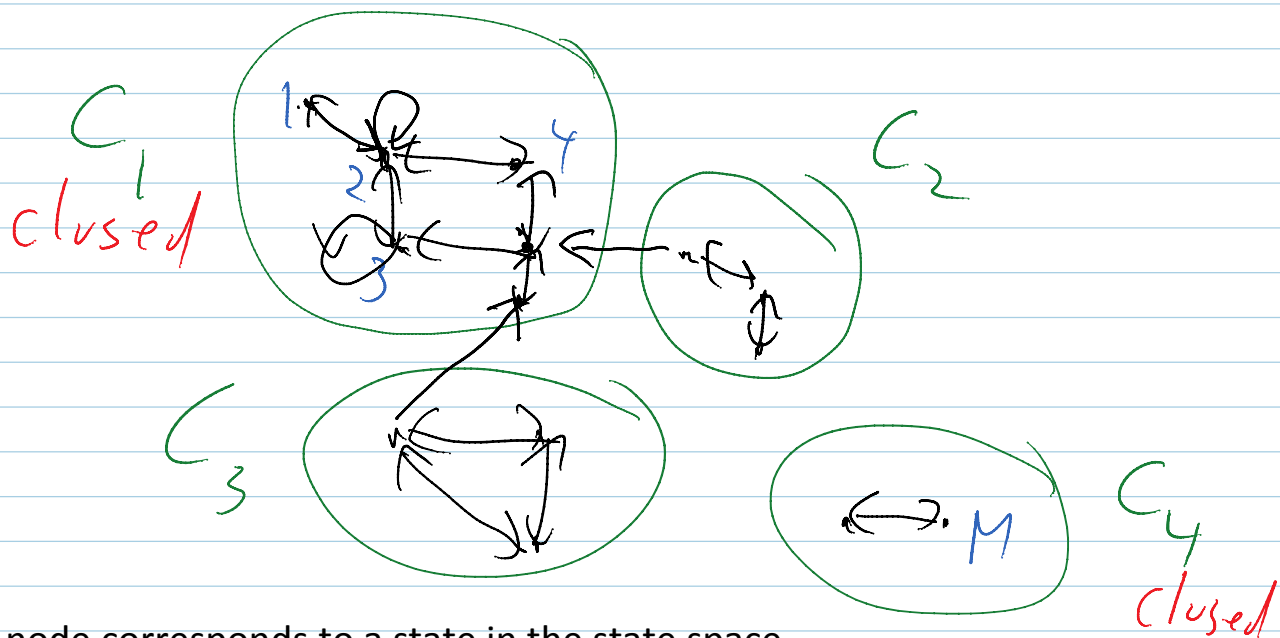
Uniqueness?

- See **Resnick Secs. 2.12-2.15** for formal proofs of statements.
- One requires a classification of the **topology of the Markov chain** in order to answer whether or not the stationary distribution is unique.
 - topology of a Markov chain simply means the properties of the graph whose adjacency matrix $A_{i,j} = 1$ if $P_{ij} > 0$ and otherwise $A_{ij} = 0$. In other words, draw a directed graph on the state space with directed edges everywhere that $P_{i,j} > 0$.

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

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

Recalling that P^n describes the matrix of probabilities to go from one state to another in exactly n epochs, this statement just means: "It is possible to go from state i to state j , and from state j to state i ." (Except state i always communicates with state i , even if $P_{ii} = 0$; choose $n = m = 0$.)



Every node corresponds to a state in the state space $\{1, 2, \dots, M\}$. A directed edge exists from node i to node j if and only if $P_{ij} > 0$.

Notationally we write $i \leftrightarrow j$ for the statement that state i communicates with state j .

We can also write a notation $i \rightarrow j$ to mean that it is possible to go from state i to state j , but not necessarily from state j to state i . (This is not enough for communication.)

Communication is an equivalence relation, and a consequence from algebra is that the state space can be decomposed into **communication classes** $\{C_r\}$ such that all states within a communication class communicate with each other, but states in different communication classes do not communicate with each other.

A communication class C_r is said to be **closed** provided there exist no pair of states $i \in C_r$, $k \notin C_r$ such that $i \rightarrow k$. In other words, there is no way to leave a closed communication class.

A Markov chain is said to be **irreducible** if it has a single communication class (meaning that all states communicate with each other.)

- Note that if one restricts a reducible Markov chain to a closed communication class, the result is an irreducible Markov chain with state space equal to that closed communication class.
 - this will be useful in applying some results for irreducible Markov chains to reducible Markov chains

The key uniqueness result for FSDT MCs is:

Any irreducible FSDT MC has a unique stationary distribution, which moreover can be expressed by the following formula:

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

where

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

is the first passage time for state j (usually called first return time if $X_0 = j$), which is a random variable.

The conditional expectation of random variables can be computed by using the conditional probabilities:

$$\mathbb{E}[T_j(1)|X_0 = j] = \sum_{n=1}^{\infty} n P(T_j(1) = n|X_0 = j)$$

The intuition behind the linkage between the stationary distribution and first return time is as follows: Suppose we have a statistically stationary Markov chain initialized by the stationary distribution π . Then for any epoch n , we have $P(X_n = j) = \pi_j$, which means that the MC will be in state j roughly a fraction π_j of the time.

So if we record the states visited by a realization of the MC with stationary distribution

$$\pi_1 = \frac{1}{2}, \quad \pi_2 = \frac{1}{12}, \quad \pi_3 = \frac{5}{12}$$

| } } | 2 | } | } } | } } 2 | } }

The connection between the stationary distribution and the average first return time is usually used as a means to compute average first return times.