

Homework 2 will be posted by tomorrow morning, due Friday, October 16 at 5 PM.

To prepare to describe the conditions under which the stationary distribution serves as a limit distribution for a FSDT MC, we need to introduce one more characterization of the probability transition matrix:

The **period** of a state i of a FSDT MC is defined as follows:

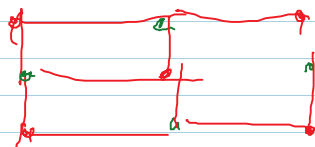
$$d(i) = \gcd\{n > 0: (P^n)_{ii} > 0\}$$

the greatest common divisor of set of integers such that state i can return to state i in exactly that many epochs.

The period of a state is a **class property**, meaning it takes the same value for all states in a communication class. A state (and therefore a communication class) is said to be **aperiodic** whenever the period of the state/class is equal to 1.

The typical situation in which a Markov chain class fails to be aperiodic is when the Markov chain is somehow forced to cycle through certain subclasses of states in a synchronized fashion; there is a theorem that says in fact this is true in a mathematical sense. (see [Lawler Ch. 1.](#))

Random walks on graphs can sometimes fail to be aperiodic when the random walker is forced to move in each epoch. A somewhat common situation is bipartite graphs (means you can decompose the graph into two subgraphs such that every edge connects nodes from one subgraph to the other subgraph). Then the period would in general be 2 for undirected graphs (and possibly larger for directed graphs).



In the last two paragraphs, subclass does not mean communication class.

Now we can make two important statement about how stationary distributions represent long-time properties of FSDT MCs:

1. Stationary distribution as a **limit distribution**

An **irreducible, aperiodic** FSDT MC will have the property:

$\lim_{n \rightarrow \infty} P(X_n = i) = \pi_i$ where π is the stationary distribution.

2. Law of Large Numbers for Markov Chains

An **irreducible** (but not necessarily aperiodic) FSDT MC has the property that:

For any function f on the state space S :

$$\lim_{N \rightarrow \infty} \frac{\sum_{n=1}^N f(X_n)}{N} = \sum_{i \in S} \pi_i f(i)$$

Some noteworthy aspects of the LLN FSDT MC:

- extends LLN from iid random variables to sequences of random variables obtained from Markov chain dynamics
- LHS is essentially a "long-run" average of the behavior of the FSDT MC, and the right hand side gives a typically computable deterministic formula for this long-run behavior
 - Think about the function f as some cost/reward associated to the state of the FSDT MC
- In fact the LLN FSDT MC expresses an **ergodic property**, meaning a relationship between a long time average and an "ensemble average" with respect to some probability distribution of states (here the stationary distribution). Ergodicity is a property that is often exploited or assumed in practice because it's useful for computational statistics, in fact in both directions:
 - If one can compute the RHS of the LLN FSDT MC, then this gives a good formula for computing long-time averages without having to do a long-time simulation. For example, some Hamiltonian systems in physics, which have known stationary distributions such as microcanonical, canonical, etc.
 - On the other hand, if the stationary distribution is hard to obtain or compute with, then one can compute the RHS by taking long-time averages as indicated on the left hand side
 - Imagine we are trying to calculate an average of a function on a large, high-dimensional state space S with respect to a probability distribution p for which it's hard to simulate random variables directly (high-dimensional, correlated, non-Gaussian): Want $\mathbb{E}f(X) = \sum_{i \in S} p_i f(i)$. There is an approach called **Markov Chain Monte Carlo (MCMC)** which makes up a Markov chain X_n to construct approximate calculations for this quantity, using the LLN for MC. The main challenges to doing so are:
 - Must construct Markov chain $\{X_n\}$ to have p as its stationary distribution.
 - ◆ the most common way to do this is to use the **Metropolis-Hastings algorithm**, which takes a proposed Markov chain and fixes it (by acceptance/rejection steps) so it has the right stationary distribution.
 - The Markov chain must have "good ergodic" properties.
 - This approach only makes sense if the Markov chain is not too expensive to simulate.

- Even though the LLN for MC guarantees eventual convergence, an important practical question is: **how long does one need to wait** until the LHS is actually close to the RHS. This is called the equilibration time or relaxation time, etc. for Markov chains, and trying to get this analytically is very hard on complex systems. But for simpler systems, it actually can be easily estimated by looking at the **second largest eigenvalue** of the probability transition matrix for reasons I'll explain later.
 - The reason this is important in practice is that usually the initial condition is not somehow natural/representative of the true system, and we rely on ergodicity/LLN to tell us that if we simulate the system long enough, the initial conditions don't matter much to the result.
 - and along these lines, one often applies a "**burn-in period**" to a Markov chain simulation to allow the initial condition to become more "natural," meaning closer to the stationary distribution, before taking statistics from the simulation
 - Also for MCMC, it's important that the Markov chain be constructed in such a way that the convergence to the stationary distribution is reasonably fast, i.e., the random moves have to be sufficiently aggressive but also not get rejected too much.

Note that Resnick calls a Markov chain "ergodic" if it is both aperiodic and irreducible, but this is a little confusing because periodicity does not interfere with ergodicity in the sense of the LLN.

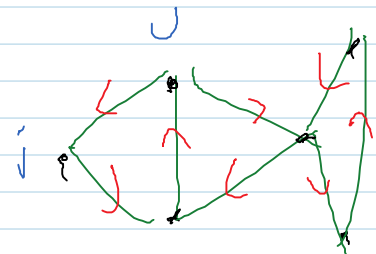
Computing stationary distributions

One way is just to use the definition of stationary distribution, i.e., find the **left eigenvector** of P with eigenvalue **1**, and then normalize it appropriately.

But there are alternatives that might be faster, particularly when the state space is large.

- **Power method**: for irreducible, aperiodic MCs: $\pi = \lim_{n \rightarrow \infty} \phi^T P^n$ where ϕ is any probability distribution vector.
 - From this perspective, we see that the second largest eigenvalue λ_2 , governs the rate of convergence, in that the difference between the LHS and RHS will have $O(\lambda_2^n)$, with $|\lambda_2| < 1$.
- $\pi = \mathbf{1}^T (I - P + \mathbf{ONE})^{-1}$ from Resnick Sec. 2.14, where $\mathbf{1}$ is a vector of 1's, and \mathbf{ONE} is a matrix of all 1's.
- Method of Kirchhof (**Haken Synergetics Secs. 4.6-4.8**): uses some graphical/diagrammatic analysis of the Markov chain to get analytical formulas; I've only seen this work well in some biochemical reaction networks
- Most important approach for obtaining **analytical solutions** for stationary distributions is to look for a **detailed balance** solution: Simply try to find a solution π that satisfies the following equation:

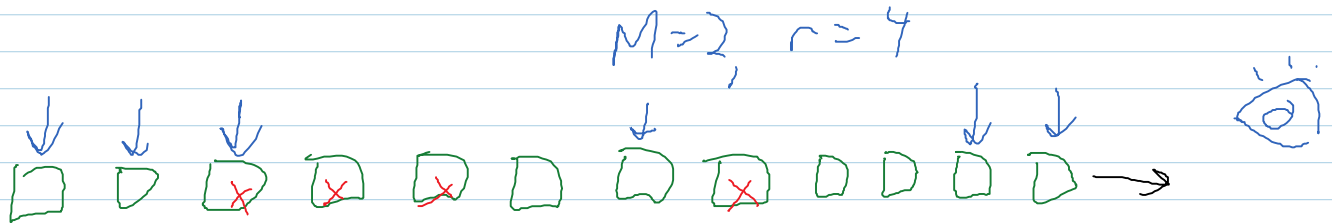
$$\pi_i P_{ij} = \pi_j P_{ji} \text{ for all } i, j \in S.$$



In general, this is overdetermined, and there will be no solution satisfying these conditions, but it's easy to check, and if such a detailed balance solution exists, then it is the stationary distribution (if irreducible).

Also, systems with time-reversal symmetry typically have detailed balance solutions, i.e. in statistical mechanics. Also random walks on graph.

Let's see how to apply the stationary distribution to answering long-run questions about a Markov chain model.



We will consider an inspection protocol for some manufactured products. The inspection protocol is as follows:

- Every manufactured product is inspected until the inspector has seen M good products in a row.
- So long as the last M inspected products are good, the inspector will only inspect every r products (i.e, skip $r - 1$ products and pass them along for shipping)
- But as soon as the inspector detects a defect, it will resume inspecting every product until it has seen M good ones in a row.

Every good inspected product, and every uninspected product is shipped, but inspected defective products are destroyed.

Practical questions (imagine trying to design M, r)

- **Cost:** fraction of manufactured items that are inspected
- **Quality:** fraction of shipped items that are defective

And it would be natural, if the company is established, to look at long-run values of these fractions.

Strategy for computing these fractions (as functions of M, r):

- develop a Markov chain model for the inspections
- solve for the stationary distribution
- Evaluate the long-run fractions of interest by using the LLN for FSDT MC

The Markov chain model not only requires a description of how the inspector behaves, but also some stochastic rule for the state of the products. We'll take the simplest model where the defect status of any product is independent of any other product, that is, the defects of the products are

determined by a Bernoulli process with "success probability" p to have a defect.