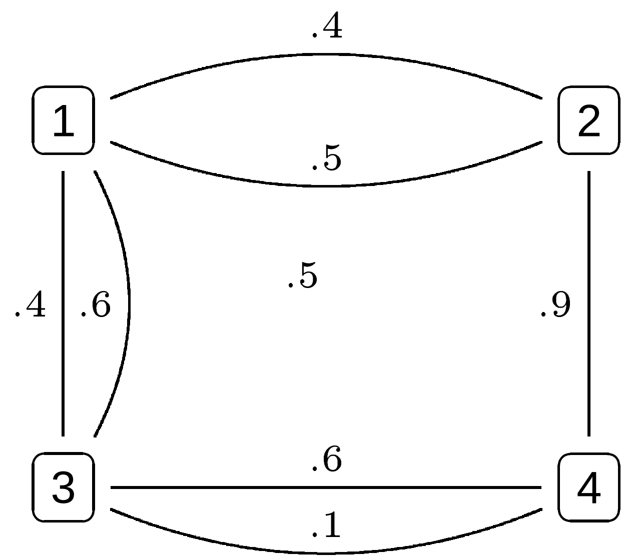
Markov Chains and Related Matters



The four nodes are called *states*.

The numbers on the arrows are called *transition probabilities*.

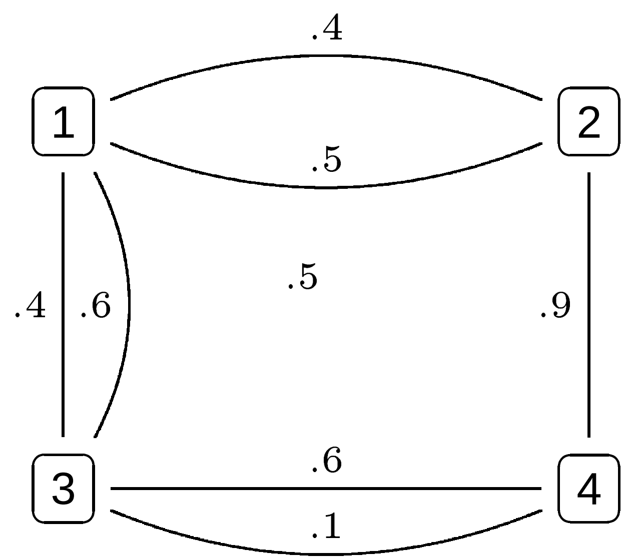
For example if we are in state 1, there is a $.6$ probability of going to state 3 and a $.4$ probability of going to state 2.

There would be no probability of staying in state 1 or going to state 4. (These are *allowed* in Markov chains, but my example just doesn't have them.)

The sum of the outgoing numbers from any state sum to 1.

These Get Applied in Many Settings

What do IU undergrads do between 6PM and midnight?



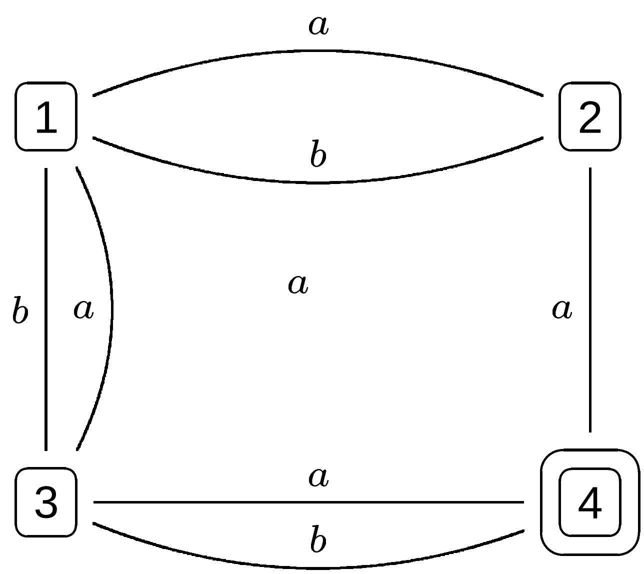
1 = study

2 = eat

3 = make phone call

4 = watch TV

Related Model: Automata



Let's declare 1 to be the *start state*.

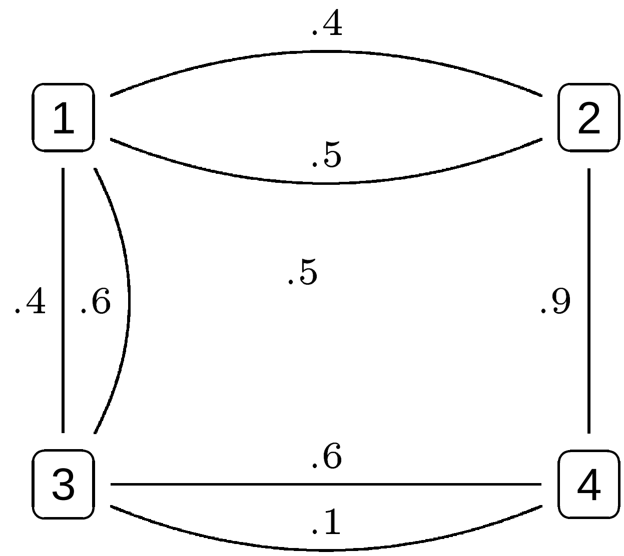
We have indicated 4 as the *accepting state*.

Then we *read in* words, following any paths of arrows we like, starting in 1 and hoping to get 4.

Some words accepted: aa, aaa, $aababababababa$.

Some words not accepted: bb, a, any word not ending in a.

Markov Chains and Related Matters



What we have here are probabilities instead of “atomic letters” like a, b, c.

We also don’t have starting or accepting state.

However, it *is* possible and useful to have these extra features “of top of” Markov chains, as we’ll see.

State Vectors

Suppose we have a large number of IU students, and we measure their "state" by a random variable X. Let's say that the value space $V(X)$ is $\{1, 2, 3, 4\}$, with

1 = study

2 = eat

3 = make phone call

4 = watch TV

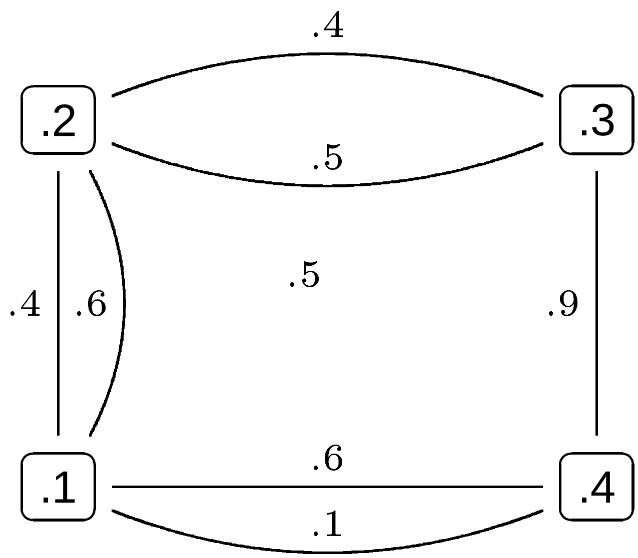
We have probabilities like *Pr*$(X = 1)$, *Pr*$(X = 2)$, etc.

These will be numbers between 0 and 1, and
$\sum_i$ *Pr*$(X = i) = 1$.

Definition A *state vector for* X is a vector $x = (x_1, \ldots, x_n)$ of numbers between 0 and 1 which sum to 1.

The Evolution of State Vectors

Suppose our original state vector x is $(.2, .3, .1, .4)$.



A good way to think about a Markov chain is that we have a large number of individuals making choices with no memory, and independent of each other.

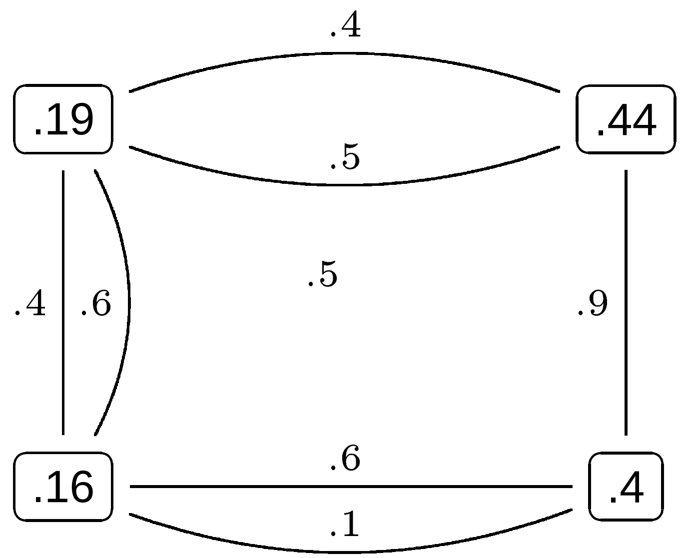
We want to go from x to x_1. How should we get X_1?

For 1, we would have $(.5)(.3) + (.4)(.1) = .15 + .04 = .19$.

The Evolution of State Vectors

We call our original state vector x_0. So x_0 is $(.2, .3, .1, .4)$.

Then we do the calculations:



We get $x_1 = (.19, .44, .31, .06)$.

Continuing, we want to get x_2, x_3, . . ., x_{3225}, . . ., x_n, ...

State Vectors and Markov Chains

We'll discuss formulas soon, but first let's remember that state vectors come from random variables.

This means that in addition to the state variables $x_0, x_1, \ldots, x_n, \ldots$ we also have random variables $X_0, X_1, \ldots, X_n, \ldots$.

These have an important independence property:

$$\textit{Pr}(X_n | X_1, \ldots, X_{n-1}) \quad = \quad \textit{Pr}(X_n | X_{n-1}). \qquad (1)$$

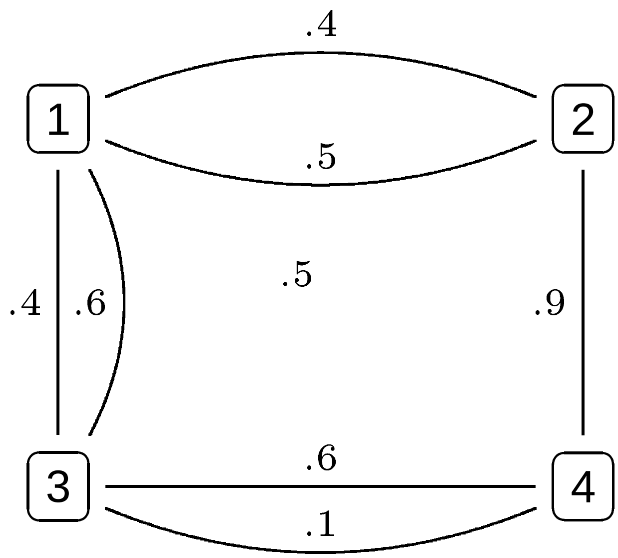
In formally, what happens at time n only depends on time $n-1$ *and not on anything further back*.

This *Markov property* in (1) formalizes the idea that we are modeling a system with *no memory*.

In sophisticated books, a Markov chain is *defined* to be the sequence of random variables, and then the picture is a *specification* of it.

Markov Chains and Related Matters

There are two ways to turn a Markov Chain into a *matrix*:



$$\begin{pmatrix} 0 & .4 & .6 & 0 \\ .5 & 0 & .5 & 0 \\ .4 & 0 & 0 & .6 \\ 0 & .9 & .1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & .5 & .4 & 0 \\ .4 & 0 & 0 & .9 \\ .6 & .5 & 0 & .1 \\ 0 & 0 & .6 & 0 \end{pmatrix}$$

Matrices, continued

$$\begin{pmatrix} 0 & .4 & .6 & 0 \\ .5 & 0 & .5 & 0 \\ .4 & 0 & 0 & .6 \\ 0 & .9 & .1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & .5 & .4 & 0 \\ .4 & 0 & 0 & .9 \\ .6 & .5 & 0 & .1 \\ 0 & 0 & .6 & 0 \end{pmatrix}$$

The first way, the rows add to 1. The second way, the *columns* do.

We always write $a_{i,j}$ for the entries in a matrix A, with the *row index first*.

On the left $p_{i,j}$ is the probability of going from state i to state j.

On the right $p_{i,j}$ is the prob of going from state j to state i.

The two matrices are *transposes* of each other.

One has to be careful in reading books and papers, because both kinds of matrices are used.

Markov Chains and Related Matters

Even more, one has to be careful in writing! Cf. Textbook, Chapter 10: both conventions seem to be in use!

To make life easier for us, we'll call P the one on the left, and P^t the one on the right.

So the rows of P add to 1.

Remember that we started with a state vector $x_0 = (.2, .3, .1, .4)$. To get x_1, we multiply matrices:

$$x_1 \quad = \quad x_0 P$$

In more detail,

$$x_1 \quad = \quad (.2, .3, .1, .4) \begin{pmatrix} 0 & .4 & .6 & 0 \\ .5 & 0 & .5 & 0 \\ .4 & 0 & 0 & .6 \\ 0 & .9 & .1 & 0 \end{pmatrix}$$

Markov Chains, continued

The general formula is $x_{n+1} = x_n P$, where the operation here is *matrix multiplication*.

We could also work with P^t, and here what we would do is to multiply on the *right* by the transpose of x_n:

$$x_{n+1}^t \quad = \quad P^t x_n^t.$$

For example, we would get x_1^t by multiplying $P^t x_0^t$:

$$\begin{pmatrix} 0 & .5 & .4 & 0 \\ .4 & 0 & 0 & .9 \\ .6 & .5 & 0 & .1 \\ 0 & 0 & .6 & 0 \end{pmatrix} \begin{pmatrix} .2 \\ .3 \\ .1 \\ .4 \end{pmatrix} \quad = \quad \begin{pmatrix} .19 \\ .44 \\ .31 \\ .06 \end{pmatrix}$$

Regular Markov Chains

Here is one of the main facts about Markov Chains.

We say that a Markov chain is *regular* if for each pair of states i and j, it is possible to go from i to j following *non-zero* arrows.

Our example chain is regular, but we'll soon see one which is not.

Then, there is a unique *long-term* state vector x^* with the following properties:

1. $x^*P = x^*$: so x^* is *stable*.
2. In terms of the transpose, $P^t(x^*)^t = (x^*)^t$.
3. No matter what the starting x_0 is, the sequence

   $$x_0, x_1, \ldots, x_n, \ldots,$$

   will *converge* to x^* in an appropriate sense.
4. We can find x^* by solving a set of linear equations.

How to find x^*

In our example, let's write x^* as (w, x, y, z). It will be easier to work with the transpose P^t and to solve

We want to solve

$$\begin{pmatrix} 0 & .5 & .4 & 0 \\ .4 & 0 & 0 & .9 \\ .6 & .5 & 0 & .1 \\ 0 & 0 & .6 & 0 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}$$

When we multiply out the left side, we get

$$\begin{pmatrix} 0w & + & .5x & + & .4y & + & z \\ .4w & + & 0x & + & 0y & + & .9z \\ .6w & + & .5x & + & 0y & + & .1z \\ 0w & + & 0x & + & .6y & + & 0z \end{pmatrix} = \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}$$

Solving it, continued

So we would have to solve

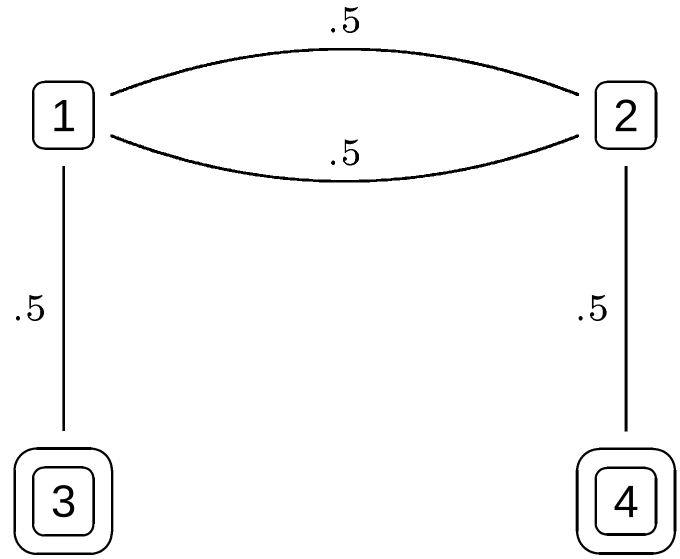
$$
\begin{array}{rcrcrcrcr}
0w & + & .5x & + & .4y & + & z & = & w \\
.4w & + & 0x & + & 0y & + & .9z & = & x \\
.6w & + & .5x & + & 0y & + & .1z & = & y \\
0w & + & 0x & + & .6y & + & 0z & = & z
\end{array}
$$

This is four linear equations in four unknowns, and it has a unique solution by our theorem.

The easiest way to get the solution is to use a program like Matlab.

A Non-Regular Chain

Flip a fain coin forever, stopping when you get a Heads. What's the probability that the first H is on an even numbered flip?



1 = an even number of T's and no H.

2 = an odd number of T's and no H.

3 = H after an an even number of all T flips

4 = H after an an odd number of all T flips

We start with $x_0 = (1, 0, 0, 0)$, and we want to know the "long term" distribution.