Absorbing Markov chains (sections 11.1 and 11.2)

Matrices of transition probabilities

Let's revisit random walk on the interval {1, 2, 3, 4} (note the change in notation: before, we used {0, 1, 2, 3}) and put it in a more general framework.

When the walker is at position i, he has some probability $p_{i,j}$ of moving to position j.

We present the numbers $p_{i,j}$ in a matrix **P** called a <u>transition matrix</u>.

```
P = {{1, 0, 0, 0}, {1 / 2, 0, 1 / 2, 0}, {0, 1 / 2, 0, 1 / 2}, {0, 0, 0, 1}}
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In some versions of *Mathematica*, the default style of presenting a list of lists looks more like this:

$$\left\{\{1, 0, 0, 0\}, \left\{\frac{1}{2}, 0, \frac{1}{2}, 0\right\}, \left\{0, \frac{1}{2}, 0, \frac{1}{2}\right\}, \{0, 0, 0, 1\}\right\}$$

If you find this happening, look in the *Mathematica* help for "Print Matrix".

```
MatrixForm[P]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This is a row-stochastic matrix: the entries in each row form a probability distribution (i.e., they are non-negative numbers that sum to 1).
Usually we will just call such a matrix stochastic.
(A square matrix that is both row-stochastic and column-stochastic is called doubly-stochastic.)

Every stochastic matrix **P** is associated with a random process that at each discrete time step is in some state, such that the probability of moving to state j at the next step is equal to $p_{i,j}$, where i is the current state.

Such a process is called a Markov chain. Sometimes we will call the states s_1, s_2, ... instead of 1, 2, ...

Note that the probability of the chain going to state j at the next time step depends ONLY on what state i the chain is in NOW, not on what states the chain visited previously.

We call **P** the transition matrix associated with the Markov chain.

Absorbing states and absorbing Markov chains

A state i is called absorbing if $p_{i,i} = 1$, that is, if the chain must stay in state i forever once it has visited that state.
Equivalently, $p_{i,j} = 0$ for all $j \neq i$.
In our random walk example, states 1 and 4 are absorbing; states 2 and 3 are not.

Say that state j is a **successor** of state i if $p_{i,j} > 0$. Write this as $i \to j$.

A Markov chain is called absorbing if every state i has a path of successors

$$i \to i' \to i'' \to \ldots$$

that eventually leads to an absorbing state.

In an absorbing Markov chain, the states that aren't absorbing are called transient.

Example: random walk on {1, 2, 3, 4}. This chain is an absorbing Markov chain. States 2 and 3 are transient.

A much bigger example is the stepping stone model (Example 11.12 in Grinstead and Snell); e.g., the states shown in Figure 11.1 and 11.2 come from an absorbing Markov chain with 2^{400} states, only 2 of which are absorbing.

At each stage, a random square S and a random neighbor T are chosen, and the color of S gets changed to

the color of T.
We do this on the torus (e.g., the four corner squares all count as neighbors), so that each square has exactly 8 neighbors.
The monochromatic states (colorings) are absorbing; the other states are transient.

Claim: For an absorbing Markov chain, the probability that the chain eventually enters an absorbing state (and stays there forever) is 1.

Proof: There exists some finite N such that every transient state can lead to an absorbing state in N or fewer steps, and there exists some positive ϵ such that, for every transient state i, the probability of arriving at an absorbing state in N or fewer steps is at least ϵ. Then, no matter where you start:

the probability of being in a transient state after N steps is at most $1 - \epsilon$;
the probability of being in a transient state after $2N$ steps is at most $(1-\epsilon)^2$;
the probability of being in a transient state after $3N$ steps is at most $(1-\epsilon)^3$; etc.
Since $(1-\epsilon)^n \to 0$ as $n \to \infty$, the probability of the chain visiting only transient states for all time is zero.

Claim: For an absorbing Markov chain, the time that it takes for the chain to arrive at some absorbing state (a random variable) has finite expected value.

Proof: We can bound the expected value by the convergent sum

$$(N-1)(1) + N(1-\epsilon) + N(1-\epsilon)^2 + \ldots$$

(Here we're using the formula

$$\begin{aligned}\mathrm{Exp}(X) = \mathrm{P}(X \geq 1) + \ldots + \mathrm{P}(X \geq N-1) \\ + \mathrm{P}(X \geq N) + \ldots + \mathrm{P}(X \geq 2N-1) \\ + \mathrm{P}(X \geq 2N) + \ldots + \mathrm{P}(X \geq 3N-1) \\ + \ldots\end{aligned}$$

where X denotes the number of steps that it takes for the chain to reach an absorbing state, rounded up to the next multiple of N.)

Note that this argument fills a hole in one of our earlier analyses of gambler's ruin.

Multiplying transition matrices

To multiply **P** by itself in *Mathematica*, use the operator "."

```
MatrixForm[P.P]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

```
MatrixForm[P.P.P]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{5}{8} & 0 & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{8} & 0 & \frac{5}{8} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

```
MatrixForm[MatrixPower[P, 3]]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{5}{8} & 0 & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{8} & 0 & \frac{5}{8} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Theorem 11.1: Let **P** be the transition matrix of a Markov chain. The ijth entry $p_{ij}^{(m)}$ of the matrix $\mathbf{P}^m$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after m steps.

Proof for the case m=1: Trivial.

Proof for the case m=2: Replace j by k and write $p_{ik}^{(2)} = \sum_{j=1}^{n} p_{ij}\ p_{jk}$.
The jth term in the RHS is equal to the probability, given that one is already at i, of going to j at the next step and to k at the step after that. Summing over j, we get the total probability of going to k in two steps.

Proof for higher cases: Left to you (same idea, more complicated notation).

Theorem 11.2: Let **P** be the transition matrix of a Markov chain, and let **u** be the probability row-vector which represents the starting distribution. Then the probability that the chain is in state i after m steps is the ith entry in the vector

$$\mathbf{u}^{(m)} = \mathbf{u}\,\mathbf{P}^m.$$

Proof: Left to you.

We'll be interested in raising **P** to ever-higher powers. E.g., for our random walk example:

```
N[MatrixPower[P, 100]]
```

$$\begin{pmatrix} 1. & 0. & 0. & 0. \\ 0.666667 & 7.88861\times 10^{-31} & 0. & 0.333333 \\ 0.333333 & 0. & 7.88861\times 10^{-31} & 0.666667 \\ 0. & 0. & 0. & 1. \end{pmatrix}$$

This tells us that if you start from state 2, you have about a .333333 chance of being in state 4 a hundred time steps later.

It appears that $\mathbf{P}^m$ converges to a limit-matrix $\mathbf{P}^\infty$ as $m\to\infty$, and *Mathematica* confirms this:

```
MatrixForm[Limit[MatrixPower[P, m], m → ∞]]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We'll prove this later when we discuss the canonical form for absorbing Markov chain matrices.

Other uses of stochastic matrices

Exercise 11.2.21 (Roberts): A city is divided into 3 areas 1, 2, and 3. It is estimated that amounts u_1, u_2, and u_3 of pollution are emitted each day from these three areas.
A fraction q_{ij} of the pollution from region i ends up the next day at region j. A fraction $q_i = 1 - \sum_j q_{ij} > 0$ goes into the atmosphere and escapes. Let $w_i^{(n)}$ be the amount of pollution in area i after n days. Show that $w_i^{(n)} = \mathbf{u} + \mathbf{uQ} + \cdots + \mathbf{uQ}^n$.

Exercise 11.2.22: The Leontief macroeconomic model.

(The matrices aren't actually stochastic, but the idea is similar.)

An important model that is governed by stochastic matrices is mass-flow. We imagine a unit of some massy, infinitely-divisible fluid, distributed over the *n* states of some Markov chain, with each site *i* starting out with *u(i)* units of fluid.

Let **u** = (*u(1)* *u(2)* *u(3)* ... *u(n)*) be the row-vector corresponding to the initial distribution of mass.

At each time-step, the fluid that is at *i* gets distributed among all the states, with a proportion of p_{ij} of the fluid at *i* going to *j* .

After one time-step, the mass-distribution vector is the row-vector **uP**;

after another time-step, the mass-distribution vector is $\mathbf{uP}^2$; etc.

Another way to prove *P* = 1/3

We saw in the first lecture that for random walk on {1,2,3,4}, the probability that a walker who starts at 2 arrives at 4 is 1/3.

Another way to prove this is with mass-flow and the center of mass.

At each time step:
All the mass at 1 stays at 1.
The mass at 2 splits evenly between 1 and 3.
The mass at 3 splits evenly between 2 and 4.
All the mass at 4 stays at 4.
So the center of mass never changes.

At the start, all of the mass is at 2.
At the end, all of the mass is at 1 or 4.
Specifically, P of the mass is at 4 and $1-P$ of the mass is at 1, so the center of mass ends up at $P(4) + (1-P)(1) = 1+3P$.

Equating $1+3P$ and 2 gives $P = 1/3$.

This argument relies implicitly on the notion of harmonic functions.

Harmonic functions

Consider an absorbing Markov chain with state space $S = \{s_1, s_2, \ldots, s_n\}$. Let f be a function defined on S with the property that

(*) $f(i) = \sum_{j \text{ in } S} p_{ij} f(j)$

for all i, or in vector form, writing f as a column vector $\mathbf{f}$,

(**) $\mathbf{f} = \mathbf{P}\mathbf{f}$.

Then f is called a harmonic function for $\mathbf{P}$. If you imagine a game in which your fortune is $f(i)$ when you are in state i, then the harmonic condition (*) or (**) means that the game is fair in the sense that your expected

fortune after one step is the same as it was before the step. (Remember the gambler whose rising and falling fortunes correspond to the position of a random walker.) Prove that when you start in a transient state *i* your expected final fortune is equal to your starting fortune $f(i)$. In other words, a fair game on a finite state space remains fair to the end.
(Proof later.)

Example: Randomwalk on {1,2,3,4}, with $f(i)=i$.

```
MatrixForm[P]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

```
MatrixForm[f = {{1}, {2}, {3}, {4}}]
```

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

```
MatrixForm[P.f]
```

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

If the mass at site i is $m(i)$, with the $m(i)'$s summing to 1, then center of mass is at
$1m(1) + 2m(2) + 3m(3) + 4m(4)$.
Let

$$\mathbf{m} = (\ m(1) \quad m(2) \quad m(3) \quad m(4)\ )$$

be the mass-distribution vector that tells how much mass is distributed at each site; then the center of mass associated with $\mathbf{m}$ is the number $\mathbf{mf}$ (the product of the row-vector $\mathbf{m}$ and the column vector $\mathbf{f}$).
Now let $\mathbf{u}$ be the initial mass distribution.
The center of mass starts out at position $\mathbf{uf}$.
One time-step later, the mass distribution is given by $\mathbf{uP}$ (the product of the row-vector $\mathbf{u}$ and the matrix $\mathbf{P}$) and so the center of mass becomes $(\mathbf{uP})\mathbf{f}$.
But $(\mathbf{uP})\mathbf{f} = \mathbf{u}(\mathbf{Pf}) = \mathbf{uf}$, which was the center mass before the mass-flow occurred.
Likewise, after the next step of flow, the mass distribu

tion is $\mathbf{uP}^2$ and the center of mass is $(\mathbf{mP}^2)\mathbf{f} = \mathbf{mPPf} = \mathbf{mPf} = \mathbf{mf}$ as before.

So, taking the limit, $(\mathbf{m}_{P}^{\infty})\mathbf{f} = \mathbf{mf}$.

There is a 2-dimensional row-eigenspace for the matrix P and the eigenvalue 1:

```
MatrixForm[{{x, 0, 0, y}}.P]
```

$(x \quad 0 \quad 0 \quad y)$

So there must be a 2-dimensional column-eigenspace for the eigenvalue 1.

We've found one column-vector, namely f:

```
MatrixForm[P.f]
```

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

What's another column-eigenvector (for the eigen-value 1), linearly independent of **f**? ...

..?..

The all-1's vector.

```
MatrixForm[P.{{1}, {1}, {1}, {1}}]
```

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

In fact, for any Markov chain, the all-1's column-vector (write it as $\mathbf{1}$) satisfies $\mathbf{P1}=\mathbf{1}$; this is just a consequence of the fact that the matrix $\mathbf{P}$ is stochastic.
If $\mathbf{Pf}=\mathbf{f}$ then any column-vector $\mathbf{v}$ that can be written as a linear combination of $\mathbf{1}$ and $\mathbf{f}$, say $\mathbf{v}=a\mathbf{1}+b\mathbf{f}$, has the property that $\mathbf{Pv}=\mathbf{v}$:
$\mathbf{Pv}=\mathbf{P}(a\mathbf{1}+b\mathbf{f})=\mathbf{P}a\mathbf{1}+\mathbf{P}b\mathbf{f}=a\mathbf{P1}+b\mathbf{Pf}=a\mathbf{1}+b\mathbf{f}=\mathbf{v}$.

The fact that the 1-eigenspace is 2-dimensional corresponds to the fact that the mass-flow system has two independent dynamically-conserved quantities: total mass and center-of-mass.

```
Eigenvalues[P]
```

$\left\{1, 1, -\frac{1}{2}, \frac{1}{2}\right\}$

Taking the harmonic functions point of view, there is a two-dimensional space of harmonic functions, spanned by the constant function $1(x)=1$ and the linear function $f(x)=x$.

A different basis for the 2-dimensional space of harmonic functions comes from the absorption probabilities.

We already saw last time that the function $h(x)$ = the probability of getting

absorbed at the right if we

start from x

is harmonic. Last time we wrote the harmonic condition as

$$h(x) = \frac{1}{2}\, h(x-1) + \frac{1}{2}\, h(x+1)$$

for x non-absorbing, but this is equivalent to

$$\mathbf{h} = \mathbf{P}\mathbf{h}.$$

We have

$$\mathbf{h} = (0\ \ P\ Q\ \ 1)^T = (0\ \ \tfrac{1}{3}\ \ \tfrac{2}{3}\ \ 1)^T$$

where superscript- T means "tranpose". Check:

```
MatrixForm[P.{{0}, {1 / 3}, {2 / 3}, {1}}]
```

$$\begin{pmatrix} 0 \\ \frac{1}{3} \\ \frac{2}{3} \\ 1 \end{pmatrix}$$

Any multiple of h is harmonic, but there are other harmonic functions, such as the one given by the column-vector $(1\ \frac{2}{3}\ \frac{1}{3}\ 0)^T$,
whose entries give the probability of getting absorbed at the left if we start from x.
These two column-vectors form a different basis for the space of harmonic functions for this 4-state Markov chain.

Advance warning: This approach works very nicely when our Markov chain has finitely many states and our vector spaces are finite-dimensional. Later we'll see that things get more complicated when there are infinitely many states. For now, just be warned that one must be careful when stepping off the path we're currently treading!

The stepping stone model

Another application of harmonic functions is to the stepping stones model.
Consider the case of 2-colorings (black vs. white) of the 20-by-20 torus. The state space is huge, but finite, so harmonic functions can be used without the cautions that we'll learn about later.

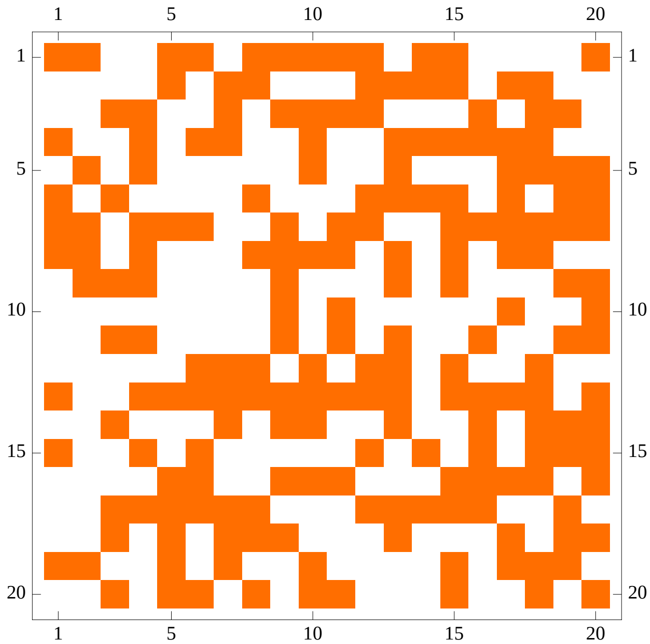
```
Size = 20
```

20

```
Board = Table[Table[RandomInteger[], {n, Size}], {m, Size}]
```

$$\begin{pmatrix}
1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}$$

```
MatrixPlot[Board]
```



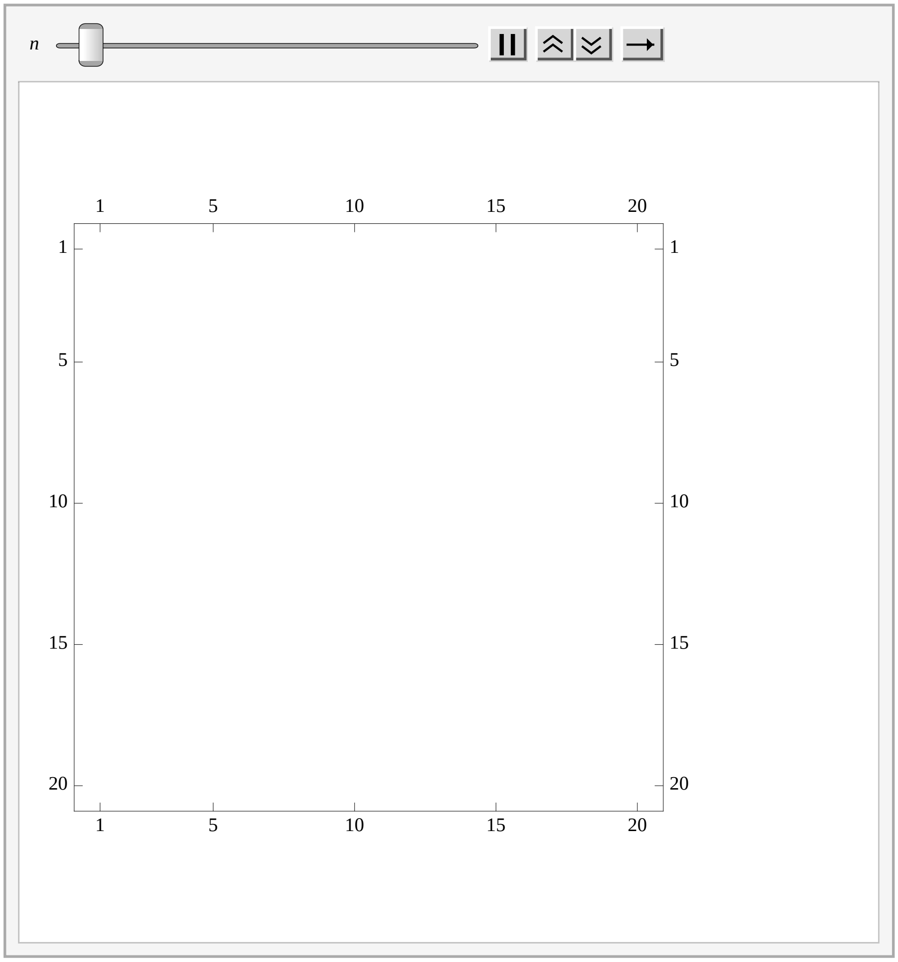

```
RandDir[] := (* random direction in grid *)
 {{1, 0}, {0, 1}, {-1, 0}, {0, -1}}[[RandomInteger[{1, 4}]]]

Wrap[x_] := (* wrap coordinates *)Which[x == 0, Size, x == Size + 1, 1, True, x]

Recolor[] := (* recolor board *)Module[{NewDir, a, b},
  NewDir = RandDir[]; a = {RandomInteger[{1, Size}], RandomInteger[{1, Size}]};
  b = {Wrap[a[[1]] + NewDir[[1]]], Wrap[a[[2]] + NewDir[[2]]]};
  Board[[b[[1]], b[[2]]]] = Board[[a[[1]], a[[2]]]]; Return[Board];]

BoardHistory := Table[Recolor[], {n, 1, 1000}];
```

```
Animate[MatrixPlot[BoardHistory[[n]]], {n, Range[1, 1000]}]
```



Given a coloring x of the 400 cells, let $f(x)$ be the proportion of white squares.

$f(x) = 1$ when all the cells are white,

$f(x) = 0$ when all the cells are black, and

$0 < f(x) < 1$ otherwise.

Claim: f is harmonic.

Proof: Instead of making S take the color of T, we could have made T take the color of S; the probability is the same. (Note: This is true because every square has the same number of neighbors as every other. That's why we made the board into a torus!) If S and T were already the same color, neither of these courses of action affects the coloring; otherwise, one of these two equally likely courses of action increases f by $\frac{1}{400}$, and the other decreases f by $\frac{1}{400}$, with an average change of 0.

More formally, if the current state is s_i, the expected value of f after one random step from s_i is a huge sum $\sum_j p_{ij}\ f(s_j)$. But we can pair up the summands, associating each index j with another index j', so that $p_{ij} = p_{ij'}$ and $f(s_j) + f(s_{j'}) = 2 f(s_i)$, so that the sum becomes $\sum_j p_{ij} f(s_i)$, which is just $f(s_i)$, which was the value of f before we took a random step.

Consequently, it may be very hard to say what sort of interface between the black and white region is likely to exist over intermediate time-scales (long enough so that some sort of law-of-large-numbers will have kicked in to smoothe out the interface, but not so long that the whole system will get sucked into an absorbing state, i.e. a monochromatic coloring), but, it is simple to figure out how likely it is that the current state will eventually become all white: it's just $f(x)$. Reasoning: Let $h(x)$ be the probability that, starting from the coloring x, the system eventually becomes all white. This function is clearly harmonic, since the equation

$h(x) = \sum_y p_{xy} h(y)$

merely encodes the fact that your probability of eventual success (in this case, success means having all cells become all white) is the weighted average of your probability of success as assessed one time step from now.

Since there are only two absorbing states, the space of row-eigenvectors for the eigenvalue 1 is only 2-dimensional; hence the space of column-eigenvectors for the eigenvalue 1 is only 2-dimensional. Since **f** and **1** are linearly independent harmonic functions, **h** must be a linear combination $\mathbf{h} = a\mathbf{f} + b\mathbf{1}$; i.e., there exist coeffi - cients a and b such that $h(x) = af(x) + b1(x) = af(x) + b$ for all x. We can solve for a and b by replacing x by the two absorbing states.

$1 = h(\text{all white}) = af(\text{all white}) + b = a + b$

$0 = h(\text{all black}) = af(\text{all black}) + b = 0 + b = b$

So $b = 0$ and $a = 1$, whence $h = f$ as claimed.

The Maximum Principle

Here's an alternative, more versatile argument for that last claim that doesn't require knowing the dimensionality of the space of harmonic functions:
Look at the function $d = h - f$, given by
$d(x) = h(x) - f(x)$ for all x.
The function d is harmonic (because it's a difference of two harmonic functions) and it vanishes at both of the absorbing states (because $h(x) = f(x) = 1$ for the all-white state and $h(x) = f(x) = 0$ for the all-black state).

Claim: A harmonic function d that vanishes at all absorbing states must vanish everywhere.

(Note: If we can prove this, then we'll have shown that $h - f = 0$, i.e., $h = f$, and we'll be done.)

Proof by contradiction: Suppose not; that is, suppose d

is non-zero somewhere.

Without loss of generality, suppose d is positive somewhere.

Let $M>0$ be the maximum value of d, and take x_0 such that $d(x_0) = M$.

Since d is harmonic, the value of d at x_0 must be a weighted average of the value of d at the successors of x_0 (remember that state y is a successor of state x if the transition probability from x to y is positive).

But all of these successors y must satisfy $d(y) \leq M$, so if even ONE successor has the property that $d(y) < M$, the weighted average of the $d(y)$'s will be less than M, which is a contradiction.

Hence every successor y of x_0 satisfies $d(y) = M$.

Now repeat the argument, using each such y in the place of x_0: We see that each successor z of each suc-

cessor y must satisfy
$d(z) = M$.
Taking this logic to its conclusion, we see that $d(x) = M$ for every state x that can be reached from x_0.
But at least one such x is an absorbing state, which by hypothesis does not satisfy $d(x) = M$; indeed, we assumed that $d(x) = 0$ whenever x is an absorbing state.
Contradiction!
Conclusion: $d(x) = 0$ for all states (transient as well as absorbing).

If this argument reminds you of a trick you learned complex analysis or electrostatics, studying continuous functions that were called "harmonic", it's not a coincidence!

In both cases, the "Maximum Principle" tells you that

a harmonic function must achieve its maximum value on the boundary of its domain.

In electrostatics, the boundary is the geometric boundary of the object that carries charge; in finite-state Markov chains, the boundary is the set of absorbing states.

Canonical form

We renumber the states so that the transient states come first. Thus, for our random walk on {1,2,3,4}, the matrix that used to be

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

becomes

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Suppose the chain has t transient states and r absorbing states. Then we can write the canonical matrix in block-form as

$$\begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where

Q is a t-by-t square matrix,

R is a non-zero t-by-r matrix,

0 is the all-zeroes r-by-t matrix, and

I is the r-by-r identity matrix.

We say such a transition matrix is in <u>canonical form</u>.

Number of visits and the fundamental matrix

Theorem 11.3: In an absorbing Markov chain, the probability that the process will be absorbed is 1 (in fact, $\mathbf{Q}^n \to 0$ exponentially as $n \to \infty$).

(Proved above.)

Consequence: $\mathbf{I} - \mathbf{Q}$ is invertible (where $\mathbf{I}$ here stands for the t-by-t identity matrix), and its inverse can be written as the convergent infinite sum $\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \ldots$. The matrix $\mathbf{N}$ is called the <u>fundamental matrix</u> for the absorbing Markov chain.

Claim: The ij-entry n_{ij} of the matrix $\mathbf{N}$ is the expected number of times the chain is in state s_j, given that it starts in state s_i. The initial state is counted (as part of "the number of times...") if $i = j$.

Proof: Fix two transient states s_i and s_j, and assume the chain starts in s_i. Let $X^{(k)}$ be a random variable that equals 1 if the chain is in state s_j after k steps, and equals 0 otherwise.

We have Prob($X^{(k)} = 1$) = $q_{ij}^{(k)}$ and Prob($X^{(k)} = 0$) = $1 - q_{ij}^{(k)}$, where $q_{ij}^{(k)}$ denotes the ijth entry of $\mathbf{Q}^k$. (Note that this works for $k = 0$ as well as $k > 0$, since $\mathbf{Q}^0 = \mathrm{I}$.) Hence E($X^{(k)}$) = $q_{ij}^{(k)}$.

The expected number of times the chain (having started in state s_i) is in state s_j in the first n steps is E($X^{(0)} + X^{(1)} + ... + X^{(n)}$) = $q_{ij}^{(0)} + q_{ij}^{(1)} + ... + q_{ij}^{(n)}$.

Sending $n \to \infty$ we have

E($X^{(0)} + X^{(1)} + ...$) = $q_{ij}^{(0)} + q_{ij}^{(1)} + ... = n_{ij}$ as claimed.

```
Q = {{0, 1/2}, {1/2, 0}};

R = {{1/2, 0}, {0, 1/2}};

N = Inverse[IdentityMatrix[2] - Q];

Set::wrsym : Symbol N is Protected. >>

FM = Inverse[IdentityMatrix[2] - Q];
```

```
MatrixForm[FM]
```

$$\begin{pmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix}$$

To see why $\frac{4}{3}$ and $\frac{2}{3}$ are correct, let x (resp. y) be the expected number of visits to 2 (resp. 3) starting from 2 (recall that 2 and 3 are transient while 1 and 4 are absorbing).

By symmetry, x is also the expected number of visits to 3 starting from 3, and y is also the expected number of visits to 2 starting from 3. So $x = 1 + (0+y)/2$ and $y = 0 + (x+0)/2$ (make sure you see where they come from!), and these equations have the unique solution

$x = \frac{4}{3}$, $y = \frac{2}{3}$.

Note that $x + y = 2$, which agrees with our earlier result that the expected number of steps until absorption (which is equal to the sum over all transient states of the expected number of visits to that states before absorption) is 2.

Theorem 11.5: Let t_i be the expected number of steps before the chain is absorbed,
given that the chain starts in state s_i, and let **t** be the column vector whose ith entry is t_i. Then $\mathbf{t} = \mathbf{N}\mathbf{c}$, where **c** is the column vector all of whose entries are 1.

Proof. If we add all the entries in the ith row of **N**, we have the expected number of times the Markov chain is in a transient state (i.e., the time until absorption), given that the chain starts in state s_i. Hence t_i is the sum of the entries in the ith row of **N**. Writing this statement in matrix form yields the theorem.