Mathematica reimbursement

Please hand in your Expense Approval forms (for purchase of copies of *Mathematica*). Today will be your last chance to do this if you want to reimbursed anytime soon.

Name of Person or Business To Be Reimbursed: *<your name>*

Date: *<today's date>*

Department: Mathematical Sciences

Remit To Address: *<your address>*

Purpose for Incurring the Expense: Purchase of software related to PI's research

Total: [make entries on separate lines for purchase price, sales tax, and total]

Signature (Person Incurring Expense): <your signature>

Ergodic Markov chains (sections 11.3 and 11.5)

Ergodicity

Definition 11.4: A Markov chain is called an ergodic chain if it is possible to go from every state to every state in a finite number of steps. (Note that no absorbing Markov chain is ergodic, aside from the trivial absorbing Markov chains with one absorbing state and no transient states.)

Example: Random walk on {1,2,3,4} with reflection on the ends (1→2 with probability 1 and 4→3 with probability 1) is ergodic.

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

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

```
MatrixForm[MatrixPower[Ref, 4]]
```

$$\begin{pmatrix} \frac{3}{8} & 0 & \frac{5}{8} & 0 \\ 0 & \frac{11}{16} & 0 & \frac{5}{16} \\ \frac{5}{16} & 0 & \frac{11}{16} & 0 \\ 0 & \frac{5}{8} & 0 & \frac{3}{8} \end{pmatrix}$$

We say a Markov chain is regular if there exists an N such that for all i and j, it is possible to go from s_i to s_j in exactly N steps; that is, the Nth power of the transition matrix **P** has all its entries positive (see Definition 11.5). Note that every regular Markov chain is ergodic.

The reflecting random walk described above is ergodic but not regular. A slightly different reflecting random walk that is regular can be obtained from our non-regular example by changing the behavior at the ends so that reflection is only partial:

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

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

```
MatrixForm[MatrixPower[ParRef, 3]]
```

$$\begin{pmatrix} \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{3}{8} & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{3}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} \end{pmatrix}$$

Recall that a function f on the state space of a Markov chain with transition matrix $\mathbf{P} = (p_{ij})$ is called harmonic if (numbering the states $1,...,n$) we have

$$f(i) = \sum_j p_{ij} f(j)$$

for all i, or equivalently, $\mathbf{Pf} = \mathbf{f}$ (where $\mathbf{f}$ denotes the column vector with entries $f(1),...,f(n)$).

Theorem: If the Markov chain with transition matrix $\mathbf{P}$ is ergodic, the only harmonic functions are the constant functions.

Proof: Let f be a harmonic function, and let M be the maximum value of f, and take x_0 such that $f(x_0) = M$. By the same reasoning that we used in the proof of the Maximum Principle for absorbing Markov chains, we can show that every successor y of x_0 satisfies $f(y) = M$, and that every successor z of every successor of x_0 satisfies $f(z) = M$, and so on. Since the chain is ergodic, every state can be reached from x_0, so we conclude that $f(s) = M$ for EVERY state s. Hence f is a constant function. □

Since the only harmonic functions are the constant functions, the only column eigenvectors for **P** with eigenvalue 1 are the multiples of the all-1's column vector.

It follows that the space of row eigenvectors with eigenvalue 1 is also 1-dimensional.

It turns out that there is a (necessarily unique) row vector **w** that is both a probability vector and a 1-eigenvector for **P**.
The components of **w** are all strictly positive.
(It is a good exercise to prove via general linear algebra that if 1 is a simple eigenvalue of the stochastic matrix **P**, i.e. if it has multiplicity 1, then the product of a row-eigenvector with eigenvalue 1 and a column-eigenvector with eigenvalue 1 is non-zero. So there is a unique row-eigenvector whose entries sum to 1. But proving that the entries are positive requires more than simple linear algebra.)

In the case where the chain is regular, **w** admits a natural interpretation: w_j is the limit of $p_{ij}^{(n)}$ as $n \to \infty$. That is, w_j is the limit, as $n \to \infty$, of the probability that a Markov chain started in state s_i that evolves via the transition matrix **P** will be in state s_j after n steps.
It is not a priori obvious that this limit exists, and indeed, for non-regular chains it does not.

Regular Markov chains

Theorem 11.7: If **P** is the transition matrix for a regular Markov chain, then as $n \to \infty$, $\mathbf{P}^n \to \mathbf{W}$ where **W** is a square matrix whose rows are all equal to the same row-vector **w**, where w is both a probability vector and a solution to $\mathbf{wP} = \mathbf{w}$. Two examples:

```
MatrixForm[N[MatrixPower[ParRef, 70]]]
```

$$\begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix}$$

```
MatrixForm[N[MatrixPower[{{1 / 2, 1 / 2}, {2 / 3, 1 / 3}}, 10]]]
```

$$\begin{pmatrix} 0.571429 & 0.428571 \\ 0.571429 & 0.428571 \end{pmatrix}$$

```
%[[1]]
```

{0.571429, 0.428571}

```
%.{{1 / 2, 1 / 2}, {2 / 3, 1 / 3}}
```

{0.571429, 0.428571}

Grinstead and Snell give two proofs in 11.4 (which we will not have time to cover).

One proof looks at how $\mathbf{P}^n\mathbf{y}$ behaves as $n \to \infty$, where **y** is an arbitrary column-vector.

The other proof looks at how $\mathbf{u}\mathbf{P}^n$ behaves as $n \to \infty$, where **u** is an arbitrary probability row-vector.

Note that the following assertions are equivalent for a non-zero row-vector w:

w is a left 1-eigenvector for **P**

w is a row 1-eigenvector for **P**

$\mathbf{wP} = \mathbf{w}$

w is a fixed vector for **P**

w is invariant under **P**

$\mathbf{wP} - \mathbf{w} = \mathbf{0}$

w is in the null-space of $\mathbf{P} - \mathbf{I}$

w is in the null-space of $\mathbf{I} - \mathbf{P}$

(where **I** is the identity matrix of the appropriate size).

When we run the Markov chain starting from an initial state governed by the probability distribution **w**, its state at each later time-step is also governed by **w**. (That is, if

$$\mathrm{Prob}(X_0 = i) = w_i \text{ for all } i,$$

with X_n denoting the state of the chain at time n, then we also have

$$\mathrm{Prob}(X_1 = i) = w_i \text{ for all } i,$$

$$\mathrm{Prob}(X_2 = i) = w_i \text{ for all } i,$$

etc.) We call this a <u>stationary Markov chain</u>, and call **w** the <u>equilibrium measure</u> or <u>stationary distribution</u>.

Theorem 11.9: Let **P** be the transition matrix for a regular Markov chain and **v** an arbitrary probability vector. Then $\mathbf{v}\mathbf{P}^n \to \mathbf{w}$ as $n \to \infty$.

The content of Theorem 11.9 is that **w** is a stable equilibrium (everything converges towards it).

Non-regular ergodic chains

If an ergodic chain is non-regular then there is a unique positive integer $d > 1$ (called the period of the chain) and an essentially unique (i.e., unique modulo cyclic relabelling) partition of the state-space of the chain into disjoint classes C_1, C_2, ..., C_d such that states in C_1 can only be followed by states in C_2, which can only be followed by states in C_3, etc., and states in C_d can only be followed by states in C_1.

There is still a unique probability measure **w** such that $\mathbf{wP}=\mathbf{w}$, but it is best thought of as $(\mathrm{w}^{(1)}+\mathrm{w}^{(2)}+...+\mathrm{w}^{(d)})/d$, where $w^{(1)},w^{(2)},...,w^{(d)}$ are probability measures such that $w_j^{(i)}$ is positive if state s_j is in class C_i and is zero otherwise, and where $\mathbf{w}^{(1)}\mathbf{P} = \mathbf{w}^{(2)}$, $\mathbf{w}^{(2)}\mathbf{P} = \mathbf{w}^{(3)}$, ..., and $\mathbf{w}^{(d)}\mathbf{P} = \mathbf{w}^{(1)}$.

For instance, our original model of reflecting random walk on {1,2,3,4} has period 2, with classes {1,3} and {2,4}; the uniform distribution on {1,2,3,4} should be viewed as the average of the uniform distribution on {1,3} and the uniform distribution on {2,4}.

The stationary distribution

In some cases, it's easy to solve for **w**.

If **P** is doubly-stochastic, then every constant row-vector is fixed under multiplication on the right by **P**. Hence the distribution that assigns probability $1/n$ to each of the n states of the chain is the unique invariant measure.
One way to see this is with mass flow: if we put 1 unit of mass on each site, each site sends out total mass 1 (the ith row-sum) and receives total mass 1 (the ith column-sum), so the mass distribution is invariant. Hence the all-1's row-vector is invariant under **P**, and scaling it by dividing by n turns it into an invariant probability vector.

Example: The matrix

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$$

satisfies $(1\ 1\ 1)\,\mathbf{P} = (1\ 1\ 1)$ so $(\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3})\ \mathbf{P} = (\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3})$.

Claim: If our Markov chain is a simple random walk on a connected graph (where vertices are connected by edges, with parallel edges and self-loops allowed, and it is possible to get from any vertex to any other by a chain of edges, and at each step we move from one vertex s_i to another by choosing a random edge incident with vertex s_i), then the vector

$$(\deg(s_1), \deg(s_2), \ldots \deg(s_n))$$

is invariant, where $\deg(s_i)$ is the number of edges incident with s_i. (Proved below.)

Example 1: Our first example of reflecting random walk on {1,2,3,4} had edges joining 1 to 2, 2 to 3, and 3 to 4. So deg(1) = 1, deg(2) = 2, deg(3) = 2, and deg(4) = 1, and (1,2,2,1) is an invariant vector. To turn it into a probability vector, divide by 6. Check:

```
w = {1 / 6, 2 / 6, 2 / 6, 1 / 6}
```

$$\left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right\}$$

```
w.Ref
```

$$\left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right\}$$

```
MatrixForm[w]
```

$$\begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}$$

```
(* If an array is not a list of lists but just a list,
Mathematica will treat it as a column-vector. *)
MatrixForm[{w}]
```

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

```
(* Compare: *)

MatrixForm[{{1 / 6, 1 / 3, 1 / 3, 1 / 6}}]
```

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

```
MatrixForm[{{1 / 6}, {1 / 3}, {1 / 3}, {1 / 6}}]
```

$$\begin{pmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}$$

Example 2: Our example of partially reflecting random walk on {1,2,3,4} had additional edges joining 1 to itself and 4 to itself. So deg(1) = 2, deg(2) = 2, deg(3) = 2, and deg(4) = 2, and (2,2,2,2) is an invariant vector. To turn it into a probability vector, divide by 8. Check:

```
{1 / 4, 1 / 4, 1 / 4, 1 / 4}.ParRef
```

$$\left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\}$$

Note also that the associated transition matrix `ParRef` is doubly stochastic:

```
ParRef
```

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

Proof of Claim: Use mass-flow. If we put mass deg(s_i) at state i for all i, and let it flow, then 1 unit of mass flows along every edge in both directions, so the total mass at each site doesn't change.

Sometimes the transition matrix **P** is sparse (many entries are zeroes), so we can find a fixed vector of **P** by solving the associated system of linear equations by hand.

Example:

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

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

```
Clear[a, b, c, d]
```

```
MatrixForm[{{a, b, c, d}}.Sparse] == MatrixForm[{{a, b, c, d}}]
```

$$\begin{pmatrix} \frac{b}{2}+d & \frac{a}{2}+\frac{c}{2} & \frac{a}{2}+\frac{c}{2} & \frac{b}{2} \end{pmatrix} = \begin{pmatrix} a & b & c & d \end{pmatrix}$$

Now do "linear algebra Sudoku" to find a fixed vector (don't worry about it being a probability vector):

$$\frac{b}{2} + d = a$$

$$\frac{a}{2} + \frac{c}{2} = b$$

$$\frac{a}{2} + \frac{c}{2} = c$$

$$\frac{b}{2} \qquad = d$$

Set $d = 1$.

$\frac{b}{2} = d$, so $b = 2$.

$\frac{b}{2} + d = a$, so $a = 2$.

$\frac{a}{2} + \frac{c}{2} = c$, so $\frac{a}{2} = \frac{c}{2}$, so $a = c$, so $c = 2$.

Consistency check: $\frac{a}{2} + \frac{c}{2} = b$.

Hence the invariant probability vector is $\left(\frac{2}{7}\ \frac{2}{7}\ \frac{2}{7}\ \frac{1}{7}\right)$.

Or you can let *Mathematica* do it for you:

```
NullSpace[Transpose[Sparse - IdentityMatrix[4]]]
```

```
(2 2 2 1)
```

The pinned stepping stone model

The stepping stone model, an example of an absorbing Markov chain, becomes an example of a regular Markov chain if we "pin" some of the sites, making at least one site permanently black and at least one site permanently white.

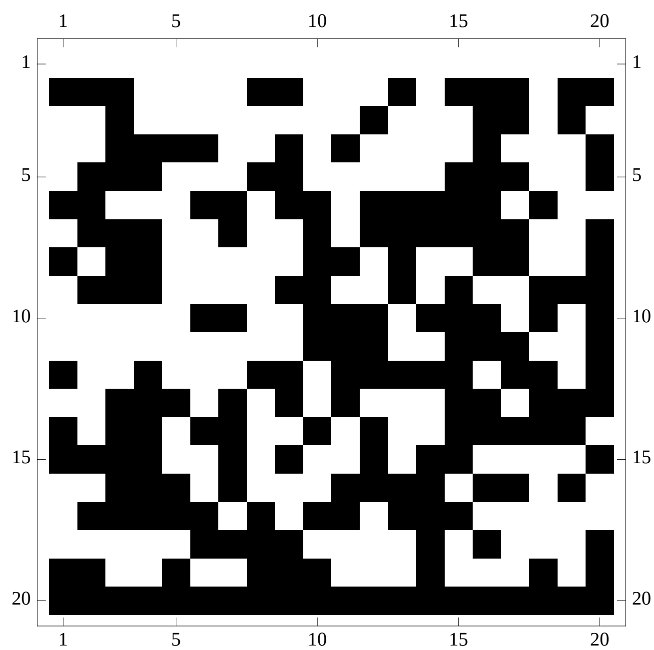
Here's some code for the pinned stepping stone model in which the top row and bottom row of a 20-by-20 grid are pinned to opposite colors:

```
Size := 20
```

```
Board =
 Table[Table[Which[i == 1, 0, i == Size, 1, True, RandomInteger[]], {j, 1, Size}], {i, 1, Size}]
```

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

```
MatrixPlot[Board, ColorFunction → "Monochrome"]
```



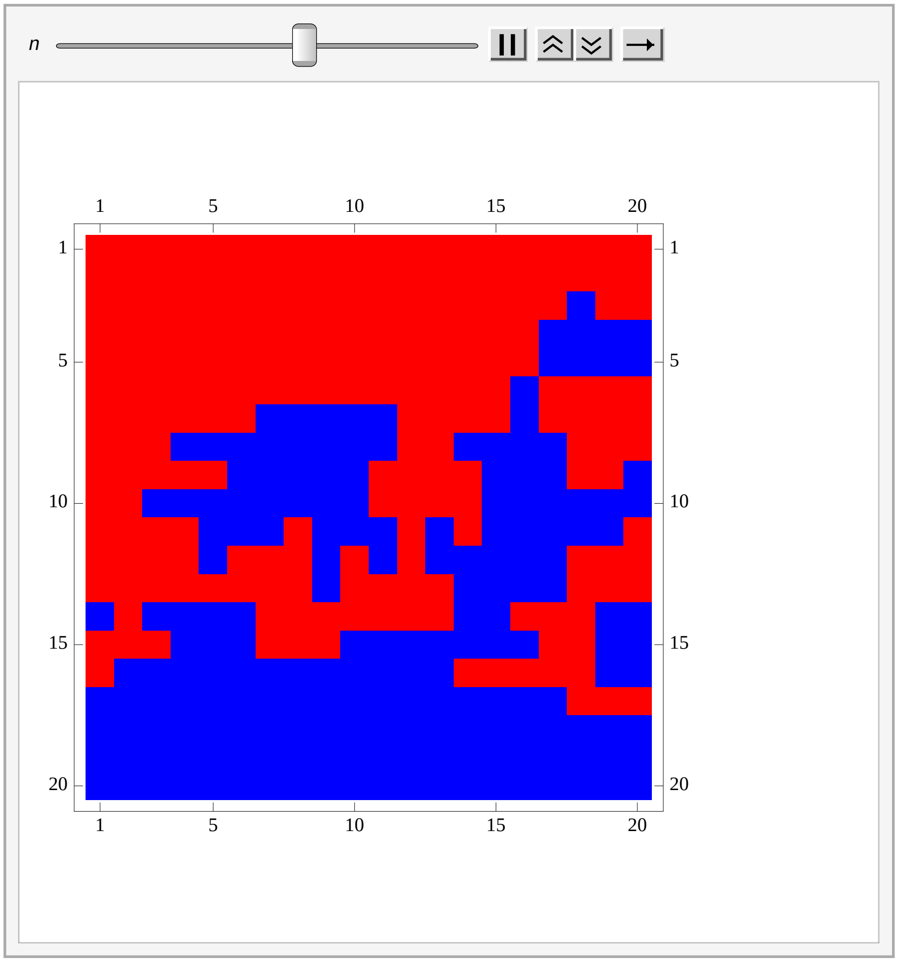

```
RandDir[] := (* random direction in grid *)
 {{1, 0}, {0, 1}, {-1, 0}, {0, -1}}[[RandomInteger[{1, 4}]]]

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[{2, Size - 1}], RandomInteger[{1, Size}]};
  b = {Wrap[a[[1]] + NewDir[[1]]], Wrap[a[[2]] + NewDir[[2]]]};
  Board[[a[[1]], a[[2]]]] = Board[[b[[1]], b[[2]]]]; Return[Board];]

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

```
Animate[MatrixPlot[BoardHistory[[n]], ColorFunction → (If[# == 0, Red, Blue] &),
  ColorFunctionScaling → False], {n, Range[1, 1000]}]
```



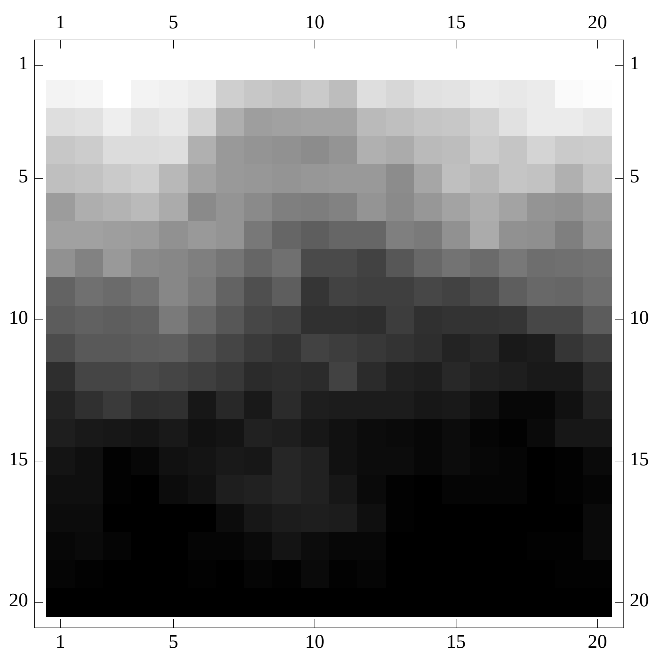
```
N[Sum[BoardHistory[[n]] / 100, {n, 1, 100}]]
```

$$\begin{pmatrix}
0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. \\
0.06 & 0.06 & 0.02 & 0.02 & 0.01 & 0.01 & 0.02 & 0.03 & 0.02 & 0.01 & 0.01 & 0.04 & 0.03 & 0.01 & 0.03 & 0.02 & 0.04 & 0.07 & 0.1 & 0.05 \\
0.12 & 0.09 & 0.09 & 0.04 & 0.03 & 0.03 & 0.06 & 0.07 & 0.05 & 0.01 & 0.02 & 0.08 & 0.05 & 0.07 & 0.09 & 0.06 & 0.07 & 0.13 & 0.12 & 0.16 \\
0.17 & 0.14 & 0.13 & 0.06 & 0.03 & 0.07 & 0.06 & 0.08 & 0.08 & 0.04 & 0.09 & 0.13 & 0.11 & 0.07 & 0.1 & 0.08 & 0.07 & 0.19 & 0.2 & 0.2 \\
0.14 & 0.13 & 0.1 & 0.03 & 0.07 & 0.08 & 0.11 & 0.09 & 0.09 & 0.07 & 0.1 & 0.21 & 0.21 & 0.2 & 0.2 & 0.22 & 0.22 & 0.29 & 0.23 & 0.18 \\
0.16 & 0.13 & 0.11 & 0.09 & 0.09 & 0.08 & 0.1 & 0.14 & 0.11 & 0.14 & 0.25 & 0.29 & 0.38 & 0.26 & 0.25 & 0.24 & 0.23 & 0.3 & 0.24 & 0.25 \\
0.33 & 0.22 & 0.15 & 0.09 & 0.05 & 0.06 & 0.11 & 0.12 & 0.11 & 0.22 & 0.25 & 0.32 & 0.32 & 0.24 & 0.29 & 0.23 & 0.23 & 0.33 & 0.38 & 0.3 \\
0.35 & 0.27 & 0.18 & 0.12 & 0.14 & 0.07 & 0.13 & 0.23 & 0.21 & 0.28 & 0.27 & 0.32 & 0.33 & 0.26 & 0.28 & 0.23 & 0.3 & 0.41 & 0.41 & 0.35 \\
0.51 & 0.33 & 0.3 & 0.22 & 0.2 & 0.15 & 0.21 & 0.25 & 0.26 & 0.33 & 0.38 & 0.33 & 0.35 & 0.28 & 0.32 & 0.28 & 0.38 & 0.41 & 0.43 & 0.48 \\
0.54 & 0.42 & 0.4 & 0.31 & 0.27 & 0.27 & 0.34 & 0.34 & 0.31 & 0.27 & 0.37 & 0.34 & 0.31 & 0.3 & 0.34 & 0.4 & 0.43 & 0.51 & 0.5 & 0.6 \\
0.64 & 0.66 & 0.56 & 0.54 & 0.45 & 0.33 & 0.42 & 0.41 & 0.28 & 0.33 & 0.38 & 0.37 & 0.34 & 0.34 & 0.38 & 0.46 & 0.52 & 0.56 & 0.61 & 0.6 \\
0.75 & 0.65 & 0.58 & 0.63 & 0.55 & 0.42 & 0.52 & 0.51 & 0.4 & 0.38 & 0.39 & 0.37 & 0.36 & 0.4 & 0.45 & 0.56 & 0.53 & 0.57 & 0.65 & 0.74 \\
0.86 & 0.84 & 0.74 & 0.76 & 0.65 & 0.52 & 0.52 & 0.56 & 0.49 & 0.46 & 0.38 & 0.34 & 0.39 & 0.48 & 0.55 & 0.66 & 0.69 & 0.79 & 0.71 & 0.84 \\
0.91 & 0.9 & 0.79 & 0.77 & 0.72 & 0.61 & 0.63 & 0.57 & 0.54 & 0.47 & 0.37 & 0.49 & 0.56 & 0.64 & 0.64 & 0.7 & 0.74 & 0.88 & 0.84 & 0.9 \\
0.95 & 0.92 & 0.86 & 0.86 & 0.76 & 0.79 & 0.69 & 0.66 & 0.68 & 0.56 & 0.5 & 0.61 & 0.69 & 0.67 & 0.71 & 0.81 & 0.82 & 0.94 & 0.94 & 0.98 \\
0.96 & 0.91 & 0.9 & 0.93 & 0.88 & 0.81 & 0.81 & 0.82 & 0.82 & 0.78 & 0.72 & 0.75 & 0.73 & 0.76 & 0.77 & 0.89 & 0.91 & 0.94 & 0.97 & 0.98 \\
0.97 & 0.98 & 0.97 & 0.88 & 0.94 & 0.92 & 0.93 & 0.9 & 0.9 & 0.87 & 0.87 & 0.89 & 0.83 & 0.84 & 0.87 & 0.92 & 0.97 & 0.98 & 0.99 & 0.99 \\
0.99 & 1. & 0.96 & 0.97 & 0.96 & 0.97 & 0.93 & 0.94 & 0.95 & 0.92 & 0.94 & 0.96 & 0.96 & 0.96 & 0.97 & 0.97 & 0.97 & 0.99 & 1. & 1. \\
1. & 1. & 1. & 0.99 & 0.98 & 0.99 & 0.97 & 0.94 & 0.98 & 0.96 & 0.97 & 0.97 & 1. & 0.99 & 0.99 & 1. & 1. & 1. & 1. & 1. \\
1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1. & 1.
\end{pmatrix}$$

```
MatrixPlot[Sum[BoardHistory[[n]] / 100, {n, 1, 100}], ColorFunction → "Grayscale"]
```

ArrayPlot::cfun :

Value of option ColorFunction −> Grayscale is not a valid color function, or a gradient ColorData entity. ≫




```
(* Why is it so slow? And what should I use instead of "Grayscale"? *)
```

Laws of large numbers

Theorem 11.12 (Weak Law of Large Numbers for Ergodic Markov Chains): Let $H_j^{(n)}$ be the proportion of times in the first n steps that an ergodic chain started from state s_i is in state s_j; i.e., it's $1/n$ times the number of visits to state s_j in the first n steps (assuming the chain started in s_i). Then for any $\epsilon > 0$,

$$\mathrm{Prob}(|H_j^{(n)} - w_j| > \epsilon) \to 0$$

regardless of i.

Strong Law of Large Numbers for Ergodic Markov Chains: With the random variable $H_j^{(n)}$ defined as above, we have $H_j^{(n)} \to w_j$ with probability 1.

Special case: If our ergodic Markov chain has the transition matrix

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

then this is just the strong law of large numbers for coin-tossing: if you toss a fair coin forever, then with probability 1, the proportion of heads up to time n approaches $\frac{1}{2}$ as $n \to \infty$.

But what does this really mean?

For simplicity, let's choose one particular value of j and suppress it from the notation, writing
$H^{(n)} \to w$.
Recall that each $H^{(n)}$ is a random variable, that is, a function $H^{(n)}(\omega)$ for ω in Ω.
Recall that $H^{(n)}(\omega) \to w$ is defined as the assertion

For all $\epsilon > 0$
there exists m such that
for all $n \geq m$,
$| H^{(n)}(\omega) - w | < \epsilon$,

which is equivalent to

For all $k \geq 1$
there exists m such that
for all $n \geq m$,
$| H^{(n)}(\omega) - w | < 1/k$

(we'll see in a minute why we want to restrict to ϵ's that are reciprocals of integers). Hence we can write the event $\{\omega: H^{(n)}(\omega) \to w\}$ as

$\bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{\omega: | H^{(n)}(\omega) - w | < 1/k \}$.

So how does probability theory work with infinite unions and infinite intersections?

"Fact": If $E_1, E_2, \ldots$ are events in Ω,

(1) $\text{Prob}(\bigcap_{n=1}^{\infty} E_n) = \lim_{N \to \infty} \text{Prob}(\bigcap_{n=1}^{N} E_n)$

and

(2) $\text{Prob}(\bigcup_{n=1}^{\infty} E_n) = \lim_{N \to \infty} \text{Prob}(\bigcup_{n=1}^{N} E_n)$.

(I call it a "Fact" with quotation marks because in a sense it's part of the way we DEFINE the probabili - ties of events like these. Then one needs a theorem that says that using these formulas can never lead to a contradiction.)

In a sense we've already encountered such trouble - some, "infinitary" events in earlier lectures. For instance, when talking about absorbing Markov chains we showed that

$E(X^{(0)} + X^{(1)} + ... + X^{(n)}) = q_{ij}^{(0)} + q_{ij}^{(1)} + ... + q_{ij}^{(n)}$.

and deduced

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

Just as it takes work to say how the probability of an event like $H^{(n)}(\omega) \to w$ should be defined, it takes work to say how the expected value of a random variable like $X^{(0)} + X^{(1)} + ...$ (an infinite sum of random variables) should be defined.

Note that for some ω's, the sum $X^{(0)} + X^{(1)} + ...$ may be infinite; but as long as the set of such ω's has probability 0, we won't worry about it.

Example: Toss a coin until it comes up Tails. This is a Markov chain with transition matrix

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

Let $X(\omega)$ be the time until the first occurrence of Tails. $X(\omega)$ is undefined if ω is the outcome

Heads, Heads, Heads, ...

but this outcome has probability $(\frac{1}{2})(\frac{1}{2})(\frac{1}{2})... = 0$.

Why did we replace $\bigcap_{\epsilon>0}$ by $\bigcap_{k=1}^{\infty}$?
Because we want to have a COUNTABLE intersection so that we can apply (1).

To see how bad things can get when we use uncountably many events at once, note that the event Ω (a set of probability 1) can be written as the union of events $\{\omega\}$, each of which has probability 0. So the rule "The probability of the union of disjoint events is the sum of the probabilities of the individual events" definitely doesn't always work when we've got more than a countable infinity of events to deal with.

But when there are only countably many, life is good: a countable union of events of
probability zero still has probability zero.

We saw an example of this in the first lecture, when we set up a correspondence between real numbers in [0,1] and infinite sequences of Heads and Tails. The correspondence breaks down for dyadic rationals, i.e. rational numbers of the form $k / 2^n$, since each of these numbers has two binary representations and hence corresponds to two differences coin-toss sequences; but since the number of dyadic rationals is countably infinite, it has Lebesgue measure 0, and the number of coin-tosses ending in infinitely many Heads or infinitely many Tails is likewise countable, so it has probability 0; so if we look at where the correspondence between real numbers and coin-toss experiments fails, we're dealing with a leftover set of reals of measure 0 that fails to correspond to a set of coin-toss outcomes with probability 0. For purposes of computing probabilities, expectations, etc., sets of measure or probability 0 can be safely ignored.

Probabilists say "For a set of ω of probability 1, $H^{(n)}(\omega) \to w$" or "For almost every ω, $H^{(n)}(\omega) \to w$" or "$H^{(n)}(\omega) \to w$ almost surely", often abbreviating "almost every" with "a.e." and "almost surely" with "a.s.", to mean "outside of a set of ω's of probability 0".

Central Limit Theorem for Markov Chains: When n is large, the distribution of the random variable $S_j^{(n)} = \mathrm{n}\mathrm{H}_j^{(n)}$ (the number of times state j occurs in the first n steps of the process) is increasingly well-approximated by a Gaussian with mean $n w_j$ and variance $\sigma_j^2 n$ (for a suitable constant $\sigma_j > 0$ whose precise value we will not compute), in the following sense: for all $r < s$,

$$\mathrm{Prob}\left[r < \frac{S_j^{(n)} - n w_j}{\sigma_j \sqrt{n}} < s\right]$$

converges to

$$\frac{1}{\sqrt{2\pi}} \int_r^s e^{-x^2/2}\, dx$$

as $n \to \infty$.

Hence $H_j^{(n)}$ is increasingly well-approximated by a Gaussian with mean w_j and variance σ_j / n, in the sense that for all $r < s$,

$$\mathrm{Prob}\left[r < \frac{H_j^{(n)} - w_j}{\sigma_j \big/ \sqrt{n}} < s\right]$$

converges to that same integral.

If we try to use the random quantity $H_j^{(n)}$ as an estimate of the non-random but unknown quantity w_j, we expect error on the order of C/sqrt(n) for some C.

First passage and recurrence times

Suppose $\mathbf{P}$ is the stochastic matrix of transition probabilities for an ergodic Markov chain, and $\mathbf{w}$ is the unique stationary measure in vector form, satisfying the three conditions

$\mathbf{w} > \mathbf{0}$ (that is, each component of $\mathbf{w}$ is positive),
$\sum_i w_i = 1$ (that is, $\mathbf{w}$ is a probability vector), and
$\mathbf{w}\mathbf{P} = \mathbf{w}$ (that is, $\mathbf{w}$ is invariant under $\mathbf{P}$).

Note that the second condition can also be written as $\mathbf{w}\mathbf{1} = 1$, where $\mathbf{1}$ on the left side of the equation is the all-1's column vector (not to be confused with the scalar 1 on the right side of the equation).

Let $\mathbf{W}$ be the square matrix each of whose rows is $\mathbf{w}$. Then $\mathbf{w}\mathbf{P} = \mathbf{w}$ implies $\mathbf{W}\mathbf{P} = \mathbf{W}$, and also $\mathbf{W}\mathbf{P}^n = \mathbf{W}$ for all $n \geq 1$.

```
P = {{1 / 2, 1 / 2}, {1 / 3, 2 / 3}}
```

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

```
Eigenvalues[P]
```

$\left\{1, \frac{1}{6}\right\}$

```
Eigenvectors[P]
```

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

```
(* What Mathematica is trying to give us here is a list of vectors forming a basis
 for the column eigenspaces.  These vectors are {1,1} and {-3/2,1} construed
 as column vectors.  Mathematica unhelpfully displays this list of vectors
 as a matrix whose rows correspond to the desired column vectors! Check: *)
```

```
{{1 / 2, 1 / 2}, {1 / 3, 2 / 3}}.{1, 1}
```

$\{1, 1\}$

```
{{1 / 2, 1 / 2}, {1 / 3, 2 / 3}}.{-3 / 2, 1}
```

$\left\{-\frac{1}{4}, \frac{1}{6}\right\}$

```
(* If that last one doesn't scream "eigenvector" at you try this: *)
```

```
{{1 / 2, 1 / 2}, {1 / 3, 2 / 3}}.{-3 / 2, 1} / {-3 / 2, 1}
```

$\left\{\frac{1}{6}, \frac{1}{6}\right\}$

(* Is there a way to force Mathematica to display
 a list of lists as a list of lists like $\left\{\{1,1\},\left\{-\frac{3}{2},1\right\}\right\}$? *)

```
(* Anyway we want the row eigenvectors not the column eigenvectors. *)
```

```
Eigenvectors[Transpose[P]]
(* this is the one we want *)
```

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

```
(* The first row is the one that corresponds to the eigenvalue 1,
and we rescale it by (2/3 + 1) to get a probability vector. *)
```

```
w = {2 / 5, 3 / 5}
```

$\left\{\frac{2}{5}, \frac{3}{5}\right\}$

```
w.P
```

$\left\{\frac{2}{5}, \frac{3}{5}\right\}$

```
W = {w, w}; MatrixForm[W]
```

$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$

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

$$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

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

$$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

Claim: $\mathbf{MW} = \mathbf{W}$ for every stochastic matrix $\mathbf{M}$.

Proof: Every column $\mathbf{c}$ of $\mathbf{W}$ is a multiple of the all 1's vector and hence satisfies $\mathbf{Mc} = \mathbf{c}$.

Consequences: $\mathbf{PW} = \mathbf{W}$ and $\mathbf{WW} = \mathbf{W}$.

Claim: $(\mathbf{P}-\mathbf{W})^n = \mathbf{P}^n - \mathbf{W}$.

Proof: By induction. It's clearly true for n=1, and if it's true for n, then

$(\mathbf{P}-\mathbf{W})^{n+1} = (\mathbf{P}-\mathbf{W})(\mathbf{P}-\mathbf{W})^n = (\mathbf{P}-\mathbf{W})(\mathbf{P}^n-\mathbf{W})$

$= \mathbf{P}^{n+1} - \mathbf{WP}^n - \mathbf{PW} + \mathbf{WW} = \mathbf{P}^{n+1} - \mathbf{W} - \mathbf{W} + \mathbf{W}$

$= \mathbf{P}^{n+1} - \mathbf{W}$, as was to be proved.

The first passage time from i to j is a random variable, namely, the time it takes for a chain started in state i to first reach state j. (If $i = j$, the first passage time is taken to be 0.)

The mean first passage time is the expected value of the first passage time, and is denoted by $m_{i,j}$ or m_{ij}. This is always finite (since we are assuming in this part of the course that our Markov chain has finitely many states).

One way to compute the mean first passage time is to turn j into an absorbing state (by giving the transition $j \to j$ probability 1 and every other transition $j \to k$ probability 0) and then compute the expected absorption time starting from i.

Example: Random walk on {1,2,3,4} with complete reflection at the ends (i.e., 1→2 with probability 1 and 4→3 with probability 1).
What is $m_{2,1}$ (the expected time from 2 to 1)?

Let $h(x)$ be the expected time until absorption at 1. We have

$h(1) = 0$, $h(2) = a$, $h(3) = b$, and $h(4) = c$.

Familiar considerations lead us to the equations

$a = 1 + (0+b)/2$,

$b = 1 + (a+c)/2$,

$c = 1 + b$.

```
Solve[{a == 1 + (0 + b) / 2, b == 1 + (a + c) / 2, c == 1 + b}, {a, b, c}]
```

$\{\{a \to 5,\ b \to 8,\ c \to 9\}\}$

Trick: Note that the expected time until absorption for random walk on {1,2,3,4} with absorption at 1 and reflection at 4 is the same as the expected time until absorption for random walk on {1,2,3,4,5,6,7} with absorption at 1 and 7. (A random walk on {1,2,3,4,5,6,7} turns into a random walk on {1,2,3,4} if we collapse 5 to 3, 6 to 2, and 7 to 1.) So we can reduce this problem to ordinary gambler's ruin and use the formulas you found for problem 11.2.26: $a = 5 \times 1$, $b = 4 \times 2$, $c = 3 \times 3$.

Here's the **Q**-matrix for the 4-state chain, if we make the state 1 absorbing:

```
Q = {{0, 1 / 2, 0}, {1 / 2, 0, 1 / 2}, {0, 1, 0}}; MatrixForm[Q]
```

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

```
Inverse[IdentityMatrix[3] - Q]
```

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

```
%.{{1}, {1}, {1}}
```

$$\begin{pmatrix} 5 \\ 8 \\ 9 \end{pmatrix}$$

If an ergodic chain is started in state i, the return time or recurrence time is the time it takes for the chain to return to state i. (This should not be confused with the first passage time from i to itself, which is always 0.) The mean recurrence time is the expected value of the recurrence time, and is denoted by r_i.

This mean recurrence time is finite when the Markov chain has finitely many states.

One way to see this is to show that

(11.4) $\quad r_i = 1 + \sum_k p_{ik}\, m_{ki}$

(hint: where are you after 1 step?) and then use the fact that every mean first passage time is finite.

A helpful companion to the above formula is

(11.2) $\quad m_{ij} = 1 + \sum_k p_{ik}\, m_{kj}$

for $i \neq j$.

(Note that for both (11.4), when $k=j$ the term $p_{ik}\, m_{kj}$ vanishes since $m_{jj} = 0$. Likewise for (11.2), when $k=i$ the term $p_{ik}\, m_{ki}$ vanishes since $m_{jj} = 0$.)

We will use these formulas to solve for r_i and the m_{ij}'s, using matrix equations.

Mean first passage matrix and mean recurrence matrix

Let the <u>mean first passage matrix</u> **M** be the matrix whose ijth entry is m_{ij} for all i,j (recall that $m_{ii} = 0$ for all i, by definition); let the <u>mean recurrence matrix</u> **D** be the matrix whose iith entry is r_i for all i and whose off-diagonal entries are all 0; and let **C** be the matrix of all 1's. Then the equations

$$m_{ii} = 1 + \sum_k p_{ik} m_{ki} - r_i$$

(both sides equal zero!) and

$$m_{ij} = 1 + \sum_k p_{ik} m_{kj}$$

are embodied in the matrix equation

$$\mathbf{M} = \mathbf{C} + \mathbf{PM} - \mathbf{D} .$$

To solve, write it as

(11.6) $$(\mathbf{I} - \mathbf{P})\, \mathbf{M} = \mathbf{C} - \mathbf{D}.$$

Theorem 11.15: For an ergodic Markov chain, the mean recurrence time for state s_i is $r_i = 1/w_i$.

Proof: Multiply both sides of Equation 11.6 by **w**, using the fact that **w**=**wP**:

$$\mathbf{0} = (\mathbf{w}-\mathbf{wP})\mathbf{M} = \mathbf{w}(\mathbf{I} - \mathbf{P})\mathbf{M} = \mathbf{w}(\mathbf{C}-\mathbf{D}) = \mathbf{wC} - \mathbf{wD}.$$

Hence **wC** (a row vector all of whose entries are equal to 1) equals **wD** (a row vector whose *i*th entry is equal to $w_i r_i$), so $1 = w_i r_i$ as claimed.

With this information we can solve for **D**, and then we can find $\mathbf{M} = (\mathbf{I} - \mathbf{P})^{-1}(\mathbf{C} - \mathbf{D})$.

Or can we? ...

Is $\mathbf{I} - \mathbf{P}$ invertible, for every ergodic Markov chain?

Quite the contrary: $(\mathbf{I} - \mathbf{P})\mathbf{1} = \mathbf{1} - \mathbf{1} = \mathbf{0}$ (where $\mathbf{1}$ is the column-eigenvector for the eigenvalue 1), so $\mathbf{I} - \mathbf{P}$ has non-trivial kernel and hence is not invertible.

What to do?

The fundamental matrix for an ergodic chain

Claim: $\mathbf{I} - \mathbf{P} + \mathbf{W}$ is invertible.

Proof: Suppose $(\mathbf{I} - \mathbf{P} + \mathbf{W})\,\mathbf{x} = \mathbf{0}$. We will show that $\mathbf{x} = \mathbf{0}$. Multiplying the equation by **w** and using the fact that $\mathbf{w}\,(\mathbf{I} - \mathbf{P}) = \mathbf{0}$ and $\mathbf{w}\mathbf{W} = \mathbf{w}$, we have $0 = \mathbf{w}(\mathbf{I} - \mathbf{P} + \mathbf{W})\,\mathbf{x} = \mathbf{w}\,\mathbf{x}$. Therefore, $\mathbf{W}\,\mathbf{x} = \mathbf{0}$ and $(\mathbf{I} - \mathbf{P})\,\mathbf{x} = \mathbf{0}$. But the second of these implies that $\mathbf{x} = \mathbf{P}\mathbf{x}$, which can only happen if $\mathbf{x}$ is a constant vector (the space of harmonic functions for an ergodic Markov chain with finite state-space is 1-dimensional). Since $\mathbf{w}\mathbf{x} = 0$, and $\mathbf{w}$ has strictly positive entries, we see that $\mathbf{x} = \mathbf{0}$. This completes the proof.

We write $\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1}$ and call it the fundamental matrix of the ergodic Markov chain.

Grinstead and Snell prove that

$$m_{ij} = (z_{jj} - z_{ij})\, r_j$$

with $r_j = 1/ w_j$.

Our two-state example:

```
MatrixForm[P]
```

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

```
MatrixForm[W]
```

$$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}$$

```
Z = Inverse[IdentityMatrix[2] - P + W];
MatrixForm[Z]
```

$$\begin{pmatrix} \frac{28}{25} & -\frac{3}{25} \\ -\frac{2}{25} & \frac{27}{25} \end{pmatrix}$$

```
r = {5 / 2, 5 / 3}
```

$$\left\{\frac{5}{2}, \frac{5}{3}\right\}$$

```
(Z[[2, 2]] - Z[[1, 2]]) r[[2]]
```

2

```
(Z[[1, 1]] - Z[[2, 1]]) r[[1]]
```

3

Check: The transit time from 1 to 2 is distributed like a geometric random variable with $p = 1/2$, so its expected value is 2.
Likewise, the transit time from 2 to 1 is distributed like a geometric random variable with $p = 1/3$, so its expected value is 3.

Let's return to random walk on {1,2,3,4} and computing the mean transit time from 2 to 1 using **Z**:

```
P4 = {{0, 1, 0, 0}, {1 / 2, 0, 1 / 2, 0}, {0, 1 / 2, 0, 1 / 2}, {0, 0, 1, 0}}; MatrixForm[P4]
```

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

```
Eigenvalues[P4]
```

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

```
(* The second eigenvalue is 1 so the second eigenvector will be the one we want. *)
```

```
Eigenvectors[Transpose[P4]]
```

$$\begin{pmatrix} -1 & 2 & -2 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$$

```
(* That is the eigenvectors are
 {{-1,2,-2,1},{1,2,2,1},{1,-1,-1,1},{-1,-1,1,1}} *)
```

```
w4 = %[[2]] / 6
```

$$\left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right\}$$

```
W4 = {w4, w4, w4, w4}; MatrixForm[W4]
```

$$\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{pmatrix}$$

```
Z4 = Inverse[IdentityMatrix[4] - P4 + W4]; MatrixForm[Z4]
```

$$\begin{pmatrix} \frac{41}{36} & \frac{11}{18} & -\frac{7}{18} & -\frac{13}{36} \\ \frac{11}{36} & \frac{17}{18} & -\frac{1}{18} & -\frac{7}{36} \\ -\frac{7}{36} & -\frac{1}{18} & \frac{17}{18} & \frac{11}{36} \\ -\frac{13}{36} & -\frac{7}{18} & \frac{11}{18} & \frac{41}{36} \end{pmatrix}$$

```
(Z4[[1, 1]] - Z4[[2, 1]]) / w4[[1]]
```

5

Note: For some purposes a better definition of the fundamental matrix is

$$\mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{W})^{-1} - \mathbf{W}$$

$$= (\mathbf{I} - \mathbf{W}) + (\mathbf{P} - \mathbf{W}) + (\mathbf{P}^2 - \mathbf{W}) + \ldots$$

but we won't be exploring those sorts of applications. (With this definition, z_{ij} has a natural interpretation: it's the expected excess number of times you visit j when you start at i, as compared with starting from equilibrium.) For the homework, use the definition of **Z** given in the book.