Ergodic Markov chains

Blocks

Suppose we play a coin-toss game where I win if the coin comes up Heads three times in a row before it comes up Tails three times in a row, and you win in the reverse case.
Then we would want to model this with an 8-state Markov chain with states HHH, HHT, ..., TTT.
More generally, if we have a Markov chain with n states, we can build a derived Markov chain whose n^k states keep track of where the first Markov chain has been for the last k steps.

E.g., for $k = 2$, if we have an ergodic Markov chain with states $1,2,\ldots,n$, we can build a "2-block" version of the chain whose states are pairs (i,j) (intuition: j is the current state of the original chain and i is the previous state).

The transition probability $p_{(i,j),(i',j')}$ for this derived chain, in terms of the transition probability for the original chain, is p_{jk} if
$j = i'$ and 0 otherwise). The stationary measure of (i,j) in this chain is just $w_i\ p_{ij}$.

Likewise, if our derived Markov chain keeps track of the last three states of the original Markov chain, then the stationary measure of (i,j,k) is $w_i\ p_{ij}\ p_{jk}$.

Etc.

Reversibility

Up till now we've written the stationary distribution as the vector w with components w_1, w_2, ... but the most common notation in the probabilistic literature is to write it as a function π with values $\pi(s_1)$, $\pi(s_2)$, ...

We also see $\pi(s_i)$ written as $\pi(i)$ or π_i.

In terms of mass-flow, the distribution that has mass π_i at each vertex i is invariant under **P** if it remains the same when $\pi_i\ p_{ij}$ units of mass flow along each edge $i \to j$. We write this as the <u>balance equation</u>

$$\sum_j \pi_i\ p_{ij} = \sum_j \pi_j\ p_{ji}$$

since the LHS is the total mass flowing <u>out of</u> i
and the RHS is the total mass flowing <u>into</u> i.

One way the balance equation can hold for all i is if the <u>detailed balance</u> equation

$$\pi_i\ p_{ij} = \pi_j\ p_{ji}$$

holds for all i and j ; that is, the amount of mass flowing from i to j equals the amount of mass flowing from j to i. In this case, we say our Markov chain is <u>reversible</u>.

The detailed balance condition

$$\pi_i\ p_{ij} = \pi_j\ p_{ji} \quad \text{for all } i, j$$

should not be confused with the "self-transpose condition"

$$p_{ij} = p_{ji} \quad \text{for all } i, j$$

which says that the <u>proportion</u> of the mass at i that goes to j equals the proportion of the mass at j that goes to i .

If the self-transpose condition holds, then P is doubly stochastic, and the uniform distribution is invariant under P, so that $\pi_i = \pi_j$ for all i,j, so that the reversibil -

ity conditions holds as well.

However, most reversible Markov chains are not self-transpose, as the following claim illustrates:

Claim: Every 2-state Markov chain is reversible.

Let's check that this is true for the 2-state Markov chain with transition matrix

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

and stationary distribution

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

Check: $\frac{2}{5} \times \frac{1}{2} = \frac{3}{5} \times \frac{1}{3}$.

Why is every 2-state Markov chain reversible?

Proof 1: Direct computation.

Proof 2: Mass flow interpretation: If total mass is conserved, and there are only two sites, then the flow from 1 to 2 must equal the flow from 2 to 1.

Proof 3: Probabilistic interpretation: $\pi_i\ p_{ij}$ is the probability of seeing an i at time n followed by a j at time $n+1$, when the Markov chain is run in its steady state (with initial distribution π). The law of large numbers for Markov chains, applied to the "2-step version" of this Markov chain (whose states are pairs of states (i,j)), with transition probabilities $p_{(i,j),(j',k)}$ equal to p_{jk} if $j=j'$ and equal to 0 otherwise), tells us that if you run the chain long enough, the frequency of the two-letter-word "ij" (i.e., the state (i,j)) should approach $\pi_i\ p_{ij}$. That is, in n steps you should see about $n\,\pi_i\ p_{ij}$ occurrences of "ij". But in any block of length n consisting of 1's and 2's, the number of occurrences of "12" differs from the number of occurrences of "21" by at most 1. So

$n\,\pi_1\,p_{12}$ and $n\,\pi_2\,p_{21}$ remain close as $n\to\infty$, implying that $\pi_1 p_{12} = \pi_2 p_{21}$.

Example: Random walk on a graph is reversible. (Recall that taking a random step means choosing a random edge at the current vertex and travelling along it to its other endpoint.)
Check: I claim that a stationary measure for the walk has π_i is proportional to degree d_i of vertex i (defined as the number of edges with an endpoint at i); specifically, $\pi_i = d_i / D$, where
$D = \sum_i d_i$ is the sum of all the degrees of all the vertices (it's also equal to twice the number of edges, where each self-loop counts as half an edge). Check detailed balance:
$\pi_i \, p_{ij} = (d_i / D) \, (n_{ij} / d_i) = (d_j / D) \, (n_{ji} / d_j) = \pi_j \, p_{ji}$
where n_{ij} is the number of edges joining states i and j.

Examples of Markov chains

(adapted from chapters 3 and 4 of Levin, Peres, and Wilmer's "Markov Chains and Mixing Times")

Terminology and notation

Note that LP&W use different notation than Grinstead and Snell do:

LP&W use Ω for the set of states of the Markov chain, whereas G&S use S for the set of states and reserve Ω for the probability space whose elements ω are finite or infinite sequences of elements of S.

LP&W write $P(x, y)$ where G&S write p_{xy}.

LP&W write

$$\mu_t = \mu_0 \, \mathbf{P}^t$$

to denote the distribution after t steps (starting from the distribution μ_0 at time 0) and they write

$$\mathrm{P}_\mu(E) = \text{the probability of the event } E$$

given that $\mu_0 = \mu$

and

$E_\mu(X)$ = the expected value of the random variable X given that $\mu_0 = \mu$.

An important special case is where the starting probability distribution μ assigns probability 1 to one state, x, and probability 0 to every other state; in that case, LP&W write P_μ and E_μ as P_x and E_x, respectively. (These notations are quite standard.)

What G&S call an "ergodic" Markov chain, LP&W call an "irreducible" Markov chain.

What G&S call a first passage time, LP&W call a hitting time.

Coupon collecting

How many times, on average, do you have to roll a die until you've seen all six faces?

Let X_t be the number of different faces of the die you-'ve seen from time 1 to time t.

E.g., $X_1 = 1$, and X_2 = either 1 or 2.

If $X_t = k < 6$, then X_{t+1} = either k or $k+1$.

More precisely, if $X_t = k$, then $X_{t+1} = k$ with probability $\frac{k}{6}$ and $X_{t+1} = k+1$ with probability $\frac{6-k}{6}$, since there are k ways to roll a face you've seen before and 6-k ways to roll a face that's new to you.

So we could model this as a 7-state Markov chain with state space {0,1,2,3,4,5,6} with

p_{ij} equal to $\frac{i}{6}$ if $j = i$, $\frac{6-i}{6}$ if $j = i+1$, and

0 otherwise, and apply the Grinstead and Snell method to compute the expected time to get from the state 0 to the absorbing state 6.

But we will do it another way.

To compute the expected value of the random variable T := the first time t for which $X_t = 6$, we break it up as $T_1 + T_2 + ... + T_6$, where

T_1 = the number of rolls required to bring X up to 1,

T_2 = the number of rolls after that required to bring X up to 2,

etc. Each T_k is geometrically distributed with expected value $\frac{6}{6-k+1}$, since the chance of rolling a face you haven't seen yet is $\frac{6-k+1}{6}$ when you've seen k-1 different faces so far. So

$\mathrm{E}(T) = \sum_{k=1}^{6} E(T_k) = \sum_{k=1}^{6} \frac{6}{6-k+1} = \sum_{k=1}^{6} \frac{6}{k} = 6 \sum_{k=1}^{6} \frac{1}{k}$.

```
6 Sum[1 / k, {k, 6}]
```

$\frac{147}{10}$

```
N[%]
```

14.7

More generally, if we have a die each of whose n faces has an equal chance of landing facing up, the expected value of the time τ until all faces have been seen is

$$n \sum_{k=1}^{n} \frac{1}{k} \approx n \int_{1}^{n} \frac{1}{x}\, dx = n \ln n$$

(for large n).

Let's test this with *Mathematica*.

Note: *Mathematica* uses the "shifted" geometric distri - bution, which takes the values 0,1,2,... instead of the values 1,2,3,...; thus, the average value of

```
RandomInteger[GeometricDistribution[p]]
```

is not $\frac{1}{p}$, but $\frac{1-p}{p} = \frac{1}{p} - 1$.

```
Mean[GeometricDistribution[p]]
```

$$\frac{1}{p} - 1$$

To get an "unshifted" geometric random variable, you must add 1.

```
Sum[1 + RandomInteger[GeometricDistribution[k / 100]], {k, 100}]
```

580

```
table = Table[Sum[1 + RandomInteger[GeometricDistribution[k / 100]], {k, 100}], {1000}];
```

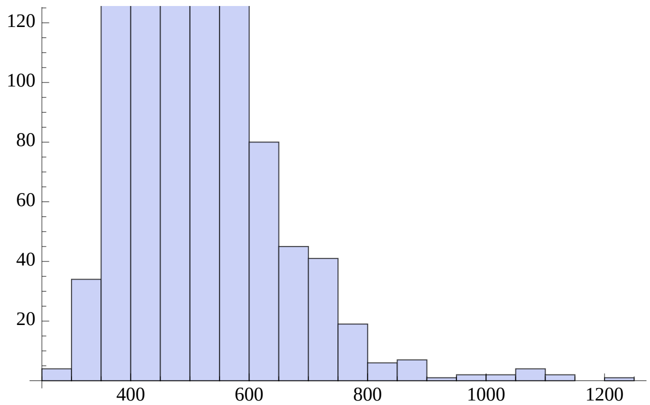
```
N[Mean[table]]
```

516.989

```
N[100 Sum[1 / k, {k, 100}]]
```

518.738

```
Histogram[table]
```



Claim: Prob($\tau > n \ln n + cn$) ≤ e^{-c}. (That is, τ is unlikely to be much more than its expected value.)

Proof: Let E_k be the event that the kth face does not appear among the first $n \ln n + cn$ rolls. Then Prob($\tau > n \ln n + cn$) = Prob($\bigcup_{k=1}^{n} E_k$) ≤ $\sum_{k=1}^{n}$Prob(E_k) = $\sum_{k=1}^{n}(1-\frac{1}{n})^{n \ln n + cn}$

$= n\,((1-\frac{1}{n})^{n})^{\ln n + c} \leq n\,(e^{-1})^{\ln n + c}$

$= n\,(e^{-\ln n})\,(e^{-c}) = n\,(1/n)\,e^{-c} = e^{-c}$.

Note: Hereafter, if I ever write "log" instead of "ln", I always mean "ln".

The Ehrenfest urn

Suppose n balls are distributed among two urns, A and B. At each move, a ball is selected at random and transferred from its current urn to the other urn. If X_t is the number of balls in urn A at time t, then X_0, X_1, ... is a Markov chain with transition probabilities

$$p_{jk} = \begin{cases} \frac{n-j}{n} & \text{if } k = j+1, \\ \frac{j}{n} & \text{if } k = j-1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that this is biased towards the middle: when X_t is bigger than $\frac{n}{2}$,

X_{t+1} tends to be smaller than X_t, and

when X_t is smaller than $\frac{n}{2}$,

X_{t+1} tends to be bigger than X_t.

Let's simulate this pseudorandomly, with $n = 100$, $X_0 = 50$:

```
RandomReal[]
```

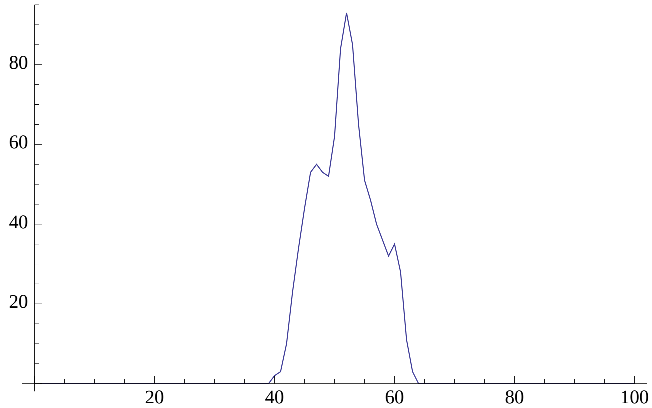
0.359172

```
c = Table[0, {100}]; k = 50; For[m = 1; k = 50, m ≤ 1000, m++, r = RandomReal[];
 If[r < (100 - k) / 100, k++, k--]; c[[k]] ++]
```

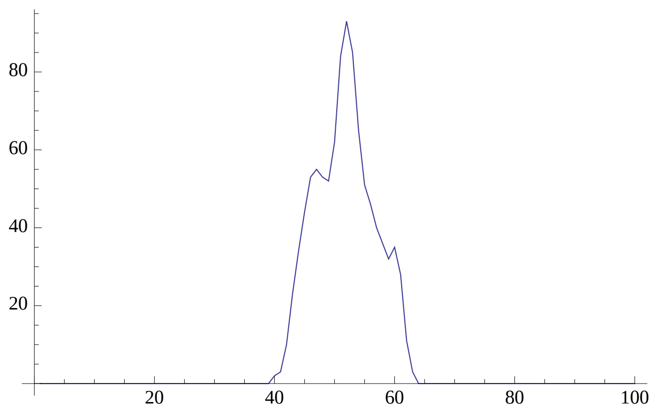
```
c
```

{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 2, 3, 10, 23, 34, 44, 53, 55, 53, 52, 62, 84, 93, 85, 65, 51, 46, 40, 36, 32, 35, 28, 11,
3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}

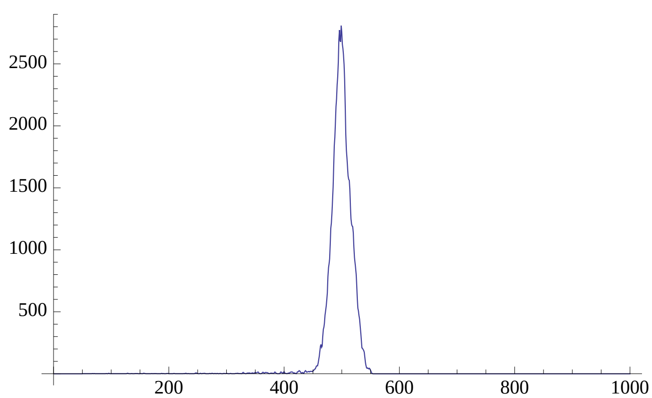
```
ListLinePlot[c]
```




```
ListLinePlot[c, PlotRange → All]
```




```
c = Table[0, {1000}]; k = 50; For[m = 1; k = 50, m ≤ 100 000, m++, r = RandomReal[];
 If[r < (1000 - k) / 1000, k++, k--]; c[[k]] ++]
ListLinePlot[c, PlotRange → All]
```



Is this approaching a Gaussian?...

On the homework, you will check directly that the binomial distribution $w_k = \binom{n}{k} / 2^n$ is an invariant probability measure for this chain.

Another way to see this is to number the balls and represent each state of the urn model by a string of n bits, where the ith bit is 1 if the ith ball is in urn A and 0 otherwise.

Then the operation of moving a random ball corresponds to the operation of flipping a random bit.

This Markov chain on bit-strings of length n is just a random walk on an n-regular graph, so the uniform distribution on bit-strings is an invariant measure for the walk.

Each bit-string has probability $1/2^n$, and $\binom{n}{k}$ of them correspond to ball-configurations with k balls in bin A, so

$$\text{Prob}[k \text{ balls in urn A}] = \binom{n}{k} / 2^n.$$

The birthday problem

Markov chain version: If people come into a room one at a time, how long do we have to wait until someone who comes in has the same birthday as someone else in the room?

Assume that there are N days in a year, and that a person is equally likely to be born on any of them.

If the first $k - 1$ people have distinct birthdays, the probability that the kth person has a different birthday from all of the first

k -1 people is $\frac{N-k+1}{N}$.

So the probability that the first n people have distinct birthdays is $p_n = \frac{N-1}{N}\frac{N-2}{N}\ldots\frac{N-(n-1)}{N}$. Approximating $\frac{N-k}{N} = 1 - \frac{k}{N}$ by $e^{-k/N}$, we get
$p_n \approx (e^{-1/N})^{1+2+\ldots+(n-1)} = (e^{-1/N})^{n(n-1)/2}$, which for
$n \approx \sqrt{N}$ gives $p_n \approx e^{-1/2} \approx 0.6$. So the value of n for which p_n crosses from $[\frac{1}{2},1]$ to $[0, \frac{1}{2}]$
is slightly larger than $\sqrt{N}$. E.g., when N = 365, the cross-over point is from n=22 to n=23.

```
N[Product[1 - (k - 1) / 365, {k, 22}]]
```

0.524305

```
N[Product[1 - (k - 1) / 365, {k, 23}]]
```

0.492703

We can model the process as an absorbing Markov chain with transient states 1,2,...,N and an absorbing state ♡, where $p_{k,k+1} = \frac{N-k}{N}$, $p_{k,\heartsuit} = \frac{k}{N}$, $p_{\heartsuit,\heartsuit} = 1$, and otherwise $p_{i,j} = 0$.

The Polya urn

Start with an urn (just one this time!) containing some black balls and some white balls (at least one of each). Choose a ball at random from those already in the urn; return the chosen ball to the urn along with another (new) ball of the same color. Repeat.

If there are a white balls and b black balls in the urn, then with probability $\frac{a}{a+b}$ a white ball will be added, and with probability $\frac{b}{a+b}$ a black balls will be added.

The (random) sequence of pairs (a,b) resulting from these choices is a Markov chain.

Example: Start with $(a,b)=(2,2)$, and run the chain for two steps. With probability $\frac{2}{4}$, a white ball will be added, and if that happens, then with probability $\frac{3}{5}$, another white ball will be added. Hence we go from (2,2) to (4,2) (in two steps) with probability $(\frac{2}{4})(\frac{3}{5}) = 0.3$. Likewise in two steps we go to (2,4) with probability 0.3, and we go to (3,3) with probability $1 - 0.3 - 0.3 = 0.4$.

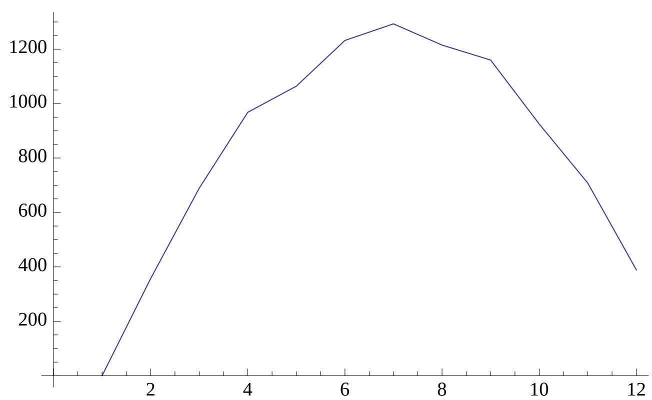
Let's run the Polya urn model for 10 steps starting from (2,2), to see where we end up; let's do this simulation repeatedly (enough times so that a smooth distribution appears).

```
c = Table[0, {12}]; For[n = 1, n ≤ 10 000, n++,
 For[{m, a, b} = {1, 2, 2}, m ≤ 10, m++, If[RandomReal[] < a / (a + b), a++, b++]]; c[[a]] ++]
```

```
c
```

{0, 357, 689, 968, 1064, 1232, 1293, 1215, 1160, 925, 708, 389}

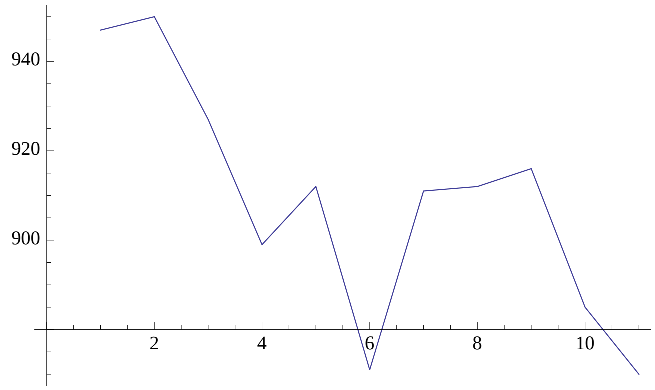
```
ListLinePlot[c, PlotRange → All]
```



What about starting from (1,1) instead of (2,2)?

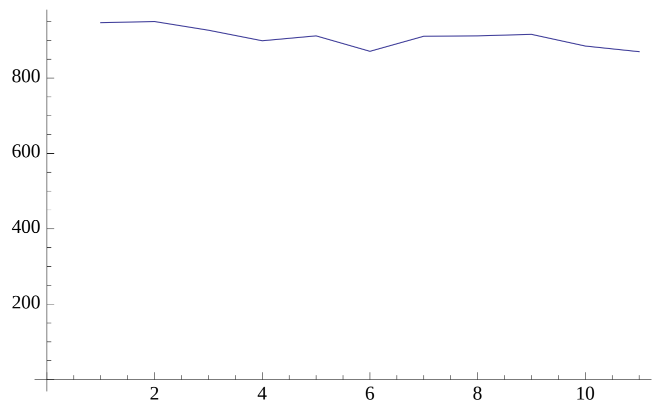
```
c = Table[0, {11}]; For[n = 1, n ≤ 10 000, n++,
 For[{m, a, b} = {1, 1, 1}, m ≤ 10, m++, If[RandomReal[] < a / (a + b), a++, b++]]; c[[a]] ++]
```

```
ListLinePlot[c, PlotRange → All]
```



The reason this graph isn't stabilizing should become clear if you look at the markings on the y-axis: the function being plotted stays within a fairly narrow range. Let's force the plot to include the x-axis.

```
ListLinePlot[c, PlotRange → All, AxesOrigin → {0, 0}]
```



It looks like the distribution of the number of white balls at time 10 is uniform on {1,...,11}!

Claim: If we start from (1,1), then the number of white balls after n steps is uniform on $\{1,...,n+1\}$. That is, if $P(a,b)$ denotes the probability of being in state (a,b) after
$a + b - 2$ steps, then $P(a,b) = \frac{1}{a+b-1}$ for all $a,b \geq 1$ (and $P(a,b) = 0$ for all other a,b).

Proof #1 (by induction on $a + b$):
The claim is trivially true for (1,1).
Suppose it's true for $(a-1,b)$ and $(a,b-1)$. Then

$$\begin{aligned}P(a,b) &= \frac{a-1}{a+b-1} P(a-1,b) + \frac{b-1}{a+b-1} P(a,b-1) \\ &= \frac{a-1}{a+b-1}\,\frac{1}{a+b-2} + \frac{b-1}{a+b-1}\,\frac{1}{a+b-2} \\ &= \frac{a+b-2}{a+b-1}\,\frac{1}{a+b-2} \\ &= \frac{1}{a+b-1}.\end{aligned}$$

Proof #2: Think about a different process, where we repeatedly add a new card at a random position in a

growing stack of cards (starting from a 1-card stack that contains just the joker).
Let a (resp. b) be 1 more than the number of cards above (resp. below) the joker.
The chance of adding the next card above the joker is $\frac{a}{a+b}$, and the chance of adding the next card below the joker is $\frac{b}{a+b}$, so the (a,b)-process for the cards is a Markov chain with the same transition probabilities as the Polya urn.
Since the stack we build in this fashion is perfectly shuffled (each permutation has the same probability as every other), the joker is as likely to be in any position as any other; hence the distribution of the (a,b) pairs is uniform.

Note that this uniformity result is very specific to starting the chain in the state (1,1). If we start it in

the state (1,2), or the state (2,1), we get slightly lop-sided distributions (try it!); but if we average the two lopsided distributions, we get something flat. Likewise for the n-ball distribution ($n > 4$)
that we get starting from (1,3), (2,2) and (3,1); these distributions are not flat, but a suitable weighted average is flat.

A variant of biased gambler's ruin

Fix some $0 < p < 1$, and take $q = 1 - p$.
Fix some positive integer n.
Consider the biased random walk on
$\{1, 2, ..., n\}$ with semiabsorbent barriers
at 1 and n that goes 1 step to the right with probability p and 1 step to the left with probability q, with the special proviso that "going 1 step to the right of n" means "staying at n" and "going 1 step to the left of 1" means "staying at 1".

Consider the mass distribution that puts mass $(p / q)^k$ at state k. You will check (in the next homework assignment) that this distribution is invariant under mass-flow.

So the stationary probability distribution has $\pi(k) = (p / q)^k / Z$, where the normalizing constant Z is $\sum_{k=1}^{n} (p / q)^k$.

(This is typical of many situations in which the explicit form of the stationary probability distribution is a nice expression times some possibly very nasty normalizing constant. In this case, the normalizing constant is easy to compute, since it's just a geometric sum; in other cases, especially those arising in statistical mechanics, the sum can be extremely hard to compute or even estimate.)

Put $n = 8$, $p = 2/5$, $q = 3/5$, $p / q = 2/3$.

```
Z = Sum[(2 / 3)^k, {k, 8}]
```

$$\frac{12610}{6561}$$

```
w = Table[((2 / 3) ^k) / Z, {k, 8}]
```

$\left\{\frac{2187}{6305}, \frac{1458}{6305}, \frac{972}{6305}, \frac{648}{6305}, \frac{432}{6305}, \frac{288}{6305}, \frac{192}{6305}, \frac{128}{6305}\right\}$

Local equilibration

```
Clear[A, B, C]
```

Clear::wrsym : Symbol C is Protected. ≫

Let A, B, C be probabilities summing to 1, and let $\mathbf{P}_1$ and $\mathbf{P}_2$ be the respective stochastic matrices

```
P1 = {{A / (A + B), B / (A + B), 0}, {A / (A + B), B / (A + B), 0}, {0, 0, 1}};
P2 = {{1, 0, 0}, {0, B / (B + C), C / (B + C)}, {0, B / (B + C), C / (B + C)}};
{MatrixForm[P1], MatrixForm[P2]}
```

$$\left\{\begin{pmatrix} \frac{A}{A+B} & \frac{B}{A+B} & 0 \\ \frac{A}{A+B} & \frac{B}{A+B} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{B}{B+C} & \frac{C}{B+C} \\ 0 & \frac{B}{B+C} & \frac{C}{B+C} \end{pmatrix}\right\}$$

Then the row-vector

```
w = {{A, B, C}}; MatrixForm[w]
```

$\begin{pmatrix} A & B & C \end{pmatrix}$

is a stationary probability vector for the matrices $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_2\mathbf{P}_1$, $\frac{1}{2}\mathbf{P}_1 + \frac{1}{2}\mathbf{P}_2$, and more generally $p\,\mathbf{P}_1 + (1\text{-}p)\,\mathbf{P}_2$ for any $0 < p < 1$.

Check:

```
w.P1
```

$\left(\frac{A^2}{A+B} + \frac{B\,A}{A+B} \quad \frac{B^2}{A+B} + \frac{A\,B}{A+B} \quad C\right)$

```
Simplify[%]
```

$\begin{pmatrix} A & B & C \end{pmatrix}$

```
Simplify[w.P2]
```

$\begin{pmatrix} A & B & C \end{pmatrix}$

Since $\mathbf{w}\mathbf{P}_1 = \mathbf{w} = \mathbf{w}\mathbf{P}_2$, the other claims follow:

$$\mathbf{w}(\mathbf{P}_1\mathbf{P}_2) = (\mathbf{w}\mathbf{P}_1)\mathbf{P}_2 = \mathbf{w}\mathbf{P}_2 = \mathbf{w}$$

$$\mathbf{w}(\mathbf{P}_2\mathbf{P}_1) = (\mathbf{w}\mathbf{P}_2)\mathbf{P}_1 = \mathbf{w}\mathbf{P}_1 = \mathbf{w}$$

$$\mathbf{w}(p\,\mathbf{P}_1 + (1-p)\,\mathbf{P}_2) = p(\mathbf{w}\mathbf{P}_1)+(1-p)(\mathbf{w}\mathbf{P}_2)$$
$$= p\mathbf{w} + (1-p)\mathbf{w} = \mathbf{w}$$

The Markov chains with transition matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are <u>not</u> ergodic, but the others ($\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_2\mathbf{P}_1$, and linear combinations of $\mathbf{P}_1$ and $\mathbf{P}_2$) <u>are</u>.

We say that the stochastic matrix $\mathbf{P}_1$ <u>locally equilibrates</u> states 1 and 2, while $\mathbf{P}_2$ locally equilibrates states 2 and 3.

In terms of mass-flow, $\mathbf{P}_1$ takes all the mass at states 1 and 2 and redistributes it between states 1 and 2 in the proportion A:B, while $\mathbf{P}_2$ takes all the mass at states 2 and 3 and redistributes it between states 2 and 3 in the proportion B:C.

The ergodic Markov chain with transition matrix $\mathbf{P}_1\mathbf{P}_2$ works by first (locally) equilibrating {1,2}, then equilibrating {2,3}, then equilibrating {1,2} again, then equilibrating {2,3} again, etc.

$\mathbf{P}_2\mathbf{P}_1$ is similar, except that it starts by equilibrating {2,3}.

$\frac{1}{2}\mathbf{P}_1 + \frac{1}{2}\mathbf{P}_2$ works by repeatedly tossing a fair coin: if the coin comes up heads, you equilibrate {1,2}, and if it comes up tails you equilibrate {2,3}.

Note that we never really used the fact that $A+B+C = 1$. So, given general positive numbers A,B,C, we have constructed several ergodic Markov chains that all preserve the probability vector ($\frac{A}{A+B+C}, \frac{B}{A+B+C}, \frac{C}{A+B+C}$).

More generally, suppose we have positive numbers A_1, ..., A_n. Let $\mathbf{P}_1$ be the transition matrix for the operation that simultaneously equilibrates {1,2}, {3,4}, ...; that is, $\mathbf{P}_1$ consists of 2-by-2 stochastic blocks, each of rank 1 (that is, the two rows of the block are the same), possibly with a 1-by-1 block consisting of just a 1 left over at the end (in the case when *n* is odd). Let $\mathbf{P}_2$ be the transition matrix for the operation that simultaneously equilibrates {2,3}, {4,5}, ... Then $\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_2\mathbf{P}_1$, and $p\,\mathbf{P}_1 + (1-p)\,\mathbf{P}_2$ (for any $0 < p < 1$) are all transition matrices for ergodic Markov chains with unique stationary probability measure ($\overline{A}_1$, ..., $\overline{A}_n$), where $\overline{A}_i = A_i / Z$ with the normalizing constant $Z = A_1 + ... + A_n$.

(This ties in with the general theme in statistical mechanics mentioned above: we often know the "weights" A_i without knowing the probabilities $\overline{A_i}$ because we can't compute Z, even though it's "just" the sum of the A_i' s.)

More generally, if we have a Markov chain with n states, and a partition Π of the state space S into disjoint subsets S_1, ..., S_k ($k > 1$), then we can do a local equilibration on each of the subsets, relative to some positive weight-function A on the state-space; this will be a non-ergodic Markov chain that preserves A.
To get an ergodic Markov chain that preserves A, we perform a succession of such Markov-updates, with a succession of different partitions Π.

In order for the resulting Markov chain to be ergodic, we need a collection $\Pi_1,\ldots,\Pi_m$ of such partitions of S, with the property that every state y of S can be reached from every other state x of S via a chain $x=x_0$, x_1, ..., $x_r=y$ such that for all k, x_k and x_{k+1} are in the same block of one of the partitions Π_i in our collection.

Given such a collection of partitions, we could just cycle through them in some fixed order; or we could at each stage roll an *m*-sided die and choose one at random.

Either way, we get an ergodic Markov chain whose unique stationary probability vector is the original specified weight-vector, scaled so that its entries sum to 1.

Later we'll see this in the context of statistical mechanics models where the weights are "Boltzmann weights", determined by the energies of the states; the rescaled weight-distribution is called the Boltzmann distribution, and this scheme for converging to the Boltzmann distribution is called heat-bath or Glauber dynamics.

Side note: If an ergodic Markov chain with transition matrix **P** has some periodicity (typically period 2), people doing simulations will often replace **P** by $\frac{1}{2}(\mathbf{I}+\mathbf{P})$, or more generally $p\mathbf{I}+(1-p)\mathbf{P}$ for some $0<p<1$, because this gets rid of periodicity (by allowing "no-ops" to occur with some positive probability) while ensuring that the stationary probability distribution **w** for **P** is still stationary for the new version of **P**. The price we pay is that our new Markov chain converges more slowly to stationarity (e.g., twice as slowly in the case of $\frac{1}{2}(\mathbf{I}+\mathbf{P})$).

Discuss HW solutions (esp. H and I)

Feedback requested!