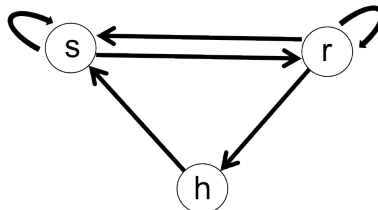


1. (**Markov chains**) Suppose that each day, the weather is either *sun*, *rain*, or *hail*. The probability of *sun* is 0.7, of *rain* is 0.2, and of *hail* is 0.1. Suppose that the weather is a Markov process, so the weather today only depends on what the weather was yesterday. If it is *hail* today, it will always be *sun* tomorrow, and if it is *sun* today, it will never *hail* tomorrow. Assume that if it is *sun* today, then it is 4 times more likely to be *sun* tomorrow than *rain*.

- (a) Draw a graphical model to represent this Markov chain.

Solution:



- (b) What is the period of each state? Is this Markov chain periodic?

Solution: All three states are aperiodic (period 1):

$$d(s) = d(r) = 1$$

$$d(h) = \gcd\{3, 4, 5, \dots\} = 1$$

The Markov chain is therefore aperiodic.

- (c) Is it irreducible? Why or why not?

Solution: This Markov chain is irreducible, since every state can be reached by every other state.

- (d) Find the matrix of transition probabilities.

Solution: From the problem statement we can obtain the entries of the transition probability matrix P below:

	<i>s</i>	<i>r</i>	<i>h</i>
<i>s</i>	0.8	0.2	0
<i>r</i>			
<i>h</i>	1	0	0

From the given stationary distribution we can then obtain p_{rh} and p_{rr} :

$$\pi_h = \pi_r \cdot p_{rh} \Rightarrow p_{rh} = 0.5$$

$$\pi_r = \pi_s \cdot p_{sr} + \pi_r \cdot p_{rr} \Rightarrow p_{rr} = 0.3.$$

Since each row needs to sum up to 1, this implies $p_{rs} = 0.2$. Therefore we get:

	<i>s</i>	<i>r</i>	<i>h</i>
<i>s</i>	0.8	0.2	0
<i>r</i>	0.2	0.3	0.5
<i>h</i>	1	0	0

2. (**Perfect phylogeny**) The table below shows a collection of 2-state full characters for $X = \{a, b, c, d, e\}$, one for each of 4 loci.

	1	2	3	4
a	0	0	1	0
b	0	0	1	0
c	0	1	1	1
d	1	0	1	1
e	1	1	0	0

Does a perfect phylogeny exist for this dataset? If yes, what is it? If no, why not?

Solution: We can use the keyword tree algorithm for perfect phylogenies presented in class.

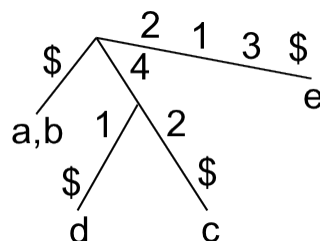
- The first step is to flip the bits in each column such that the majority allele is 1 (note: I failed to do this in the first section). The distinction between 0 and 1 is arbitrary, but it follows the convention that the ancestral state is assumed to be all 0's. Here we only need to change the 3rd character. Then the second step is to radix sort the columns:

	1	2	3	4			4	2	1	3
a	0	0	0	0		a	0	0	0	0
b	0	0	0	0		b	0	0	0	0
c	0	1	0	1	\rightarrow	c	1	1	0	0
d	1	0	0	1		d	1	0	1	0
e	1	1	1	0		e	0	1	1	1

- The third step is to encode each individual/taxon using the positions of the 1's:

a	\$
b	\$
c	4 2 \$
d	4 1 \$
e	2 1 3 \$

- The last step is to create a keyword tree using the encodings as the strings:



Since characters 1 and 2 appear more than once in the keyword tree, there does not exist a perfect phylogeny for this dataset (more than one mutation per character would be required to produce this dataset).

3. (**Ross, Ch. 4, #32**) Each of two switches is either on or off during a day. On day n , each switch will independently be on with probability

$$[1 + \text{number of on switches during day } (n - 1)]/4$$

For instance, if both switches are on during day $n - 1$, then each will independently be on during day n with probability $3/4$. What fraction of days are both switches on? What fraction are both off?

Solution: Let π_0 be the marginal probability of the state corresponding two both lights off, π_1 be the marginal probability of one light off, one light on, and π_2 be the marginal probability of both lights on. Working out the transition probabilities, we obtain:

	0	1	2
0	9/16	3/8	1/16
1	1/4	1/2	1/4
2	1/16	3/8	9/16

Then solving for the stationary distribution, we get:

$$[\pi_0, \pi_1, \pi_2] = [2/7, 3/7, 2/7].$$