Markov Chains

Gabriel T. Landi

References

- Ross, chapter 9, Sec 9.2
- Gilbert Strang, Introduction to linear algebra, Secs 8.2 (graphs and Networks) and 8.3 (Markov matrices).
- Tomé, de Oliveira, chapter 6 (Markov Chains)

Introduction

The random walk is an example of a stochastic process; ie, a process where a random variable changes with time:

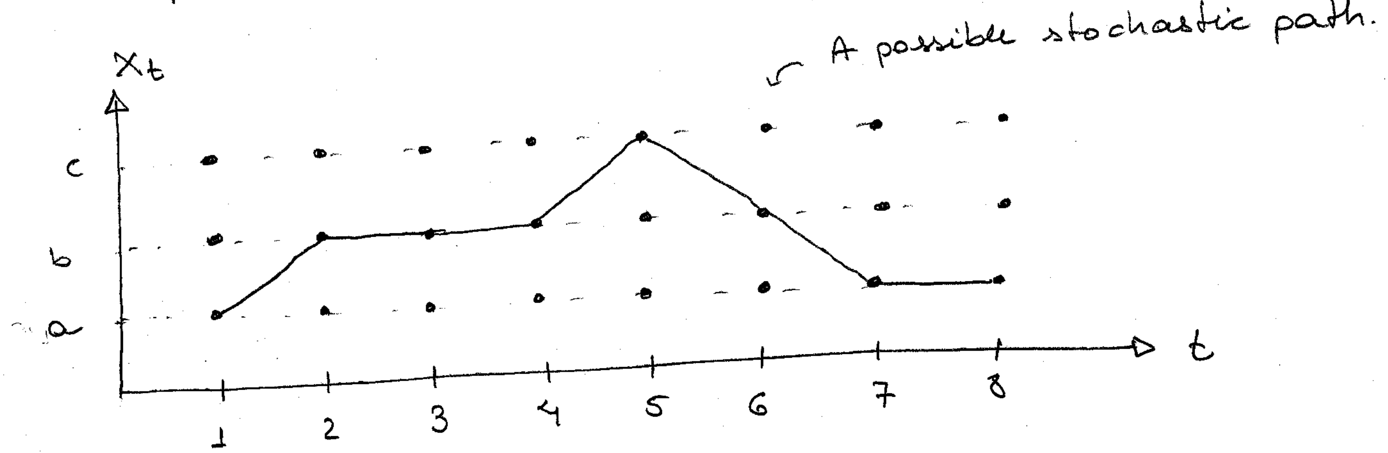
$$\boxed{X_t = \text{state of the system at time } t} \qquad (1)$$

But the random walk is somewhat limited: since you always sum independent r.v.s, the steps are always independent of the current state.

The Markov chain takes this a step further. It allows you to choose the next step based on information about your current state.

We will discuss here only the cases of discrete time and discrete space. This means that we will assume X_t takes on discrete values and that t changes in integer steps.

As the system evolves in time, X_t will take on a series of random positions. This is called a stochastic trajectory



X_t = a or b or c

The basic assumption of a Markov chain is the, so called, <u>Markov property</u>:

$$\boxed{P(X_{t+1} = m_{t+1} \mid X_t = m_t,\ X_{t-1} = m_{t-1},\ \dots,\ X_0 = m_0) = P(X_{t+1} = m_{t+1} \mid X_t = m_t)} \tag{2}$$

Think of "t" as now: this property says that <u>the future depends only on the current state</u> and not on past events. We say Markovian systems have a very short memory (like fish).

The quantities

$$\boxed{Q_{nm} = P(X_{t+1} = n \mid X_t = m) = \mathrm{Prob}(m \to n)} \tag{3}$$

are called the <u>transition probabilities</u>. They represent the probability of observing a transition from $m \to n$.

Suppose now that we know $P(X_t = m)$. We may then find $P(X_{t+1} = n)$ using the law of total probability

$$P(X_{t+1} = n) = \sum_m \overbrace{P(X_{t+1} = n \mid X_t = m)}^{Q_{nm}} P(X_t = m) \tag{4}$$

Let us switch to a more compact notation:

$$\boxed{P_n(t) = P(X_t = n)} \tag{5}$$

then Eq (4) may be written as

$$\boxed{P_n(t+1) = \sum_m Q_{nm} P_m(t)} \tag{6}$$

this is the basic formula of a Markov chain. Note that this is a <u>dynamical equation</u>. It is a recipe for <u>how to evolve your system in time</u>. (like Newton's law).

You may have noticed that Eq (6) has the structure of a <u>matrix multiplication</u>. We define a column vector

$$\vec{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \end{bmatrix} \qquad (7)$$

and a matrix Q with entries Q_{nm}. Then Eq (6) may be written as

$$\boxed{\vec{p}(t+1) = Q\,\vec{p}(t)} \qquad (8)$$

The matrix Q is called the <u>transition matrix</u>

If we start at $t=0$ with a distribution $\vec{p}(0)$, then we obtain $\vec{p}(1)$ as

$$\vec{p}(1) = Q\,\vec{p}(0)$$

and then we get $\vec{p}(2)$ as

$$\vec{p}(2) = Q\,\vec{p}(1) = Q^2\,\vec{p}(0)$$

We see that Q^2 represents the transition probabilities in a step of length 2:

$$(Q^2)_{nm} = P(X_{t+2} = n \mid X_t = m)$$

Thus, $X_0, X_2, X_4, \dots$ is also a Markov chain. Continuing with the evolution we conclude that

$$\boxed{\vec{p}(t) = Q^t\,\vec{p}(0)} \qquad (9)$$

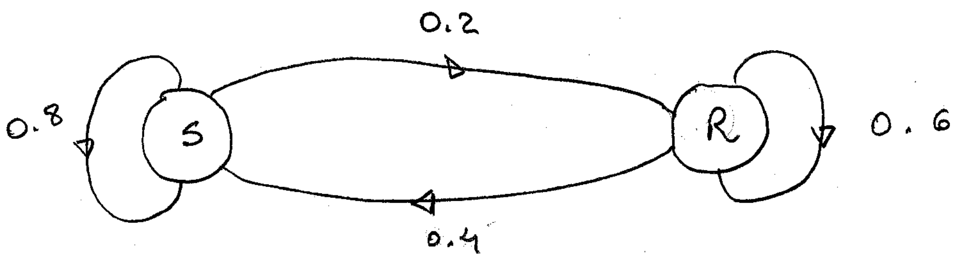
Thus, you may evolve directly from $\vec{p}(0)$ to $\vec{p}(t)$ using the matrix Q^t.

A silly example

Suppose that the weather may either be sunny or rainy. If it is sunny today, then there is a 20% chance that it will be rainy tomorrow (since water evaporates). If it is rainy today, then there is a 40% chance that it will be sunny tomorrow.

Please note that, in reality, the weather does not satisfy the Markov property (2). Its memory is certainly much better.

We may represent our silly process by a graph



(10)

This is a 2-state system. We may set "1" = "Sunny" and "2" = "Rainy". Then our system will be described by a 2-component vector

$$\vec{P}(t) = \begin{bmatrix} P(X_t = S) \\ P(X_t = R) \end{bmatrix} = \begin{bmatrix} P_1(t) \\ P_2(t) \end{bmatrix}$$

Using the law of total probability we get

$$P(X_{t+1} = S) = \underbrace{P(X_{t+1} = S \mid X_t = S)}_{0.8} P(X_t = S) + \underbrace{P(X_{t+1} = S \mid X_t = R)}_{0.4} P(X_t = R)$$

$$P(X_{t+1} = R) = \underbrace{P(X_{t+1} = R \mid X_t = S)}_{0.2} P(X_t = S) + \underbrace{P(X_{t+1} = R \mid X_t = R)}_{0.6} P(X_t = R)$$

Thus, the transition prob

$$P_1(t+1) = 0.8\, P_1(t) + 0.4\, P_2(t)$$
$$P_2(t+1) = 0.2\, P_1(t) + 0.6\, P_2(t) \qquad (11)$$

From this we immediately read off the transition matrix

$$\boxed{Q = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix}} \qquad (12)$$

Suppose that today $(t=0)$ we know its sunny. this means that

$$\vec{p}\,(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad (13a)$$

then the prob. distribution for tomorrow will be

$$\vec{p}\,(1) = Q\,\vec{p}\,(0) = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} \qquad (13b)$$

and the dist. for the day after tomorrow will be

$$\vec{p}\,(2) = Q\,\vec{p}\,(1) = \begin{bmatrix} 0.8 & 0.4 \\ 0.2 & 0.0.6 \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.72 \\ 0.28 \end{bmatrix} \qquad (13c)$$

and so on.

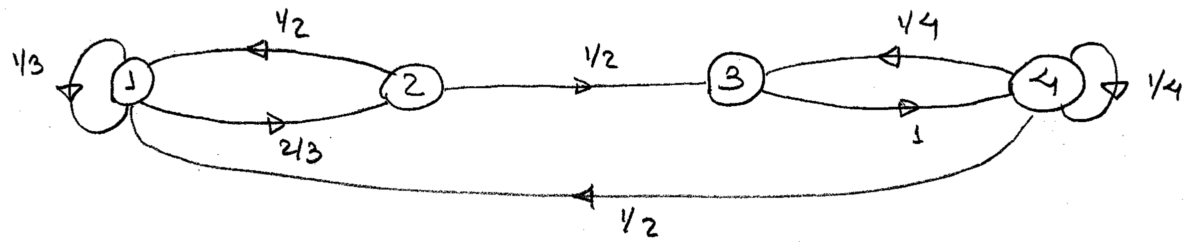
Ex: determining the transition matrix from a graph

when trying to read off the transition matrix Q from a graph, it is useful to remember that

$$Q_{mn} = P(n \to m)$$

You read the entries backwards, like arabic or japanese.

Consider, for instance



the transition matrix will be

$$Q = \begin{bmatrix} 1/3 & 1/2 & 0 & 1/2 \\ 2/3 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/4 \\ 0 & 0 & 1 & 1/4 \end{bmatrix} \tag{14}$$

Newton's law is a Markov chain

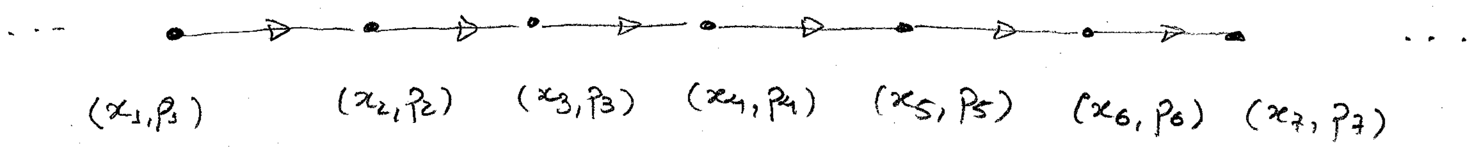
Consider Newton's 2^{nd} law for a particle in 1D:

$$\frac{dp}{dt} = F(x)$$

$$\frac{dx}{dt} = \frac{p}{m}$$

The state of the system is specified by the position and the momentum, (x, p). In terms of these two variables, Newtons law is a Markov chain: the state of the system at the next step is completely determined by the state of the system now.

Suppose we discretize time in steps of Δt. Then we may represent the evolution of $(x(t), p(t))$ as a graph:



There is, however, one big difference: at each node of the graph, there is only one edge comming in and one edge going out. If you are at (x_i, p_i) then, with 100% certainty you came from (x_{i-1}, p_{i-1}) and with 100% certainty you will go to (x_{i+1}, p_{i+1}). Newton's law is a deterministic Markov chain (trajectories in phase space never cross each other).

Stochastic matrices

The transition matrix Q satisfies some very special properties. First, its entries are always non-negative. This is obvious from the definition (3) (conditional probabilities are also probabilities!)

$$Q_{nm} = P(x_t = n \mid x_{t-1} = m) \geq 0 \qquad (15)$$

The other property Q satisfies is seen from the example in (12) and (14):

$$\boxed{\begin{array}{l}\text{The columns of } Q \text{ add up to one} \\ \displaystyle\sum_n Q_{nm} = 1 \qquad \forall m\end{array}} \qquad (16)$$

The reason why this must be true is readily seen from (6)

$$1 = \sum_n P_n(t+1) = \sum_n \Big(\sum_m Q_{nm} P_m(t) \Big)$$

$$= \sum_m \Big(\sum_n Q_{nm} \Big) P_m(t)$$

If (14) is satisfied then

$$\sum_n P_n(t+1) = \sum_m P_m(t)$$

When (14) is true, the process conserves probability: if the $P_m(t)$ are correctly normalized, then so will the $P_n(t+1)$. This may be seen in Eqs (13a) and (13b)

Any matrix Q which satisfies (15) and (16) is called a stochastic matrix. As we will soon see, these matrices satisfy a series of very important properties.

Long-time limit of our silly example: the steady-state

Let's go back to our silly example and continue the evolution in Eq (13). Here is the state of the system at further times:

$$\vec{p}(t): \quad \begin{array}{ccccc} t=0 & t=1 & t=2 & t=3 & t=4 \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix} & \begin{bmatrix} 0.72 \\ 0.28 \end{bmatrix} & \begin{bmatrix} 0.688 \\ 0.312 \end{bmatrix} & \begin{bmatrix} 0.6752 \\ 0.3248 \end{bmatrix} \end{array}$$

$$\begin{array}{cccc} t=5 & t=6 & \dots & t=1000 \\ \begin{bmatrix} 0.67008 \\ 0.32992 \end{bmatrix} & \begin{bmatrix} 0.668032 \\ 0.331968 \end{bmatrix} & & \begin{bmatrix} 0.666667 \\ 0.333333 \end{bmatrix} \end{array}$$

We see that, as time passes, the probabilities tend to "settle down" at certain values. This is what we call the steady-state distribution.

When the system reaches the steady-state, evolving it further will not change anything. So, if we let $\vec{p}^*$ denote the probabilities in the steady-state, then

$$\boxed{Q\vec{p}^* = \vec{p}^*} \tag{17}$$

This is the equation determining the steady-state. Anyone who has studied linear algebra before will recognize this as an eigenvalue/eigenvector product. The steady-state probability vector is the eigenvector of Q with eigenvalue 1. Thus, finding the steady-state is reduced to a linear algebra problem.

Let's see if this makes sense. The matrix Q is given in (12), so we compute its eigenvalues as

$$\det(Q - \lambda I) = \det \begin{bmatrix} \frac{4}{5} - \lambda & 2/5 \\ 1/5 & \frac{3}{5} - \lambda \end{bmatrix}$$

$$= \left(\tfrac{4}{5} - \lambda\right)\left(\tfrac{3}{5} - \lambda\right) - \left(\tfrac{2}{5}\right)\left(\tfrac{1}{5}\right)$$

$$= \lambda^2 - \frac{7}{5}\lambda + \frac{2}{5}$$

The eigenvalues are

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = \frac{2}{5} \tag{18}$$

The eigenvector corresponding to $\lambda_1 = 1$ is determined from

$$\frac{1}{5}\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = (1)\begin{bmatrix} a \\ b \end{bmatrix}$$

or

$$\frac{4a}{5} + \frac{2}{5}b = a$$

$$\Rightarrow \quad a = 2b \quad \Rightarrow \quad \vec{p}^* = b\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The constant b is determined from normalization: $2b + b = 1$ so $b = 1/3$. Thus we conclude that

$$\vec{p}^* = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix} \tag{19}$$

This is exactly what we found in the beginning of page 8, for $t = 1000$.

Some relevant questions concerning the steady-state

The steady-state is the most important aspect of a Markov chain. Given a certain chain, specified by a transition Matrix Q, we may ask the following questions about $\vec{p}^{\,*}$

(1) Does it exist?

(2) Is it unique?

(3) Will the chain ever reach $\vec{p}^{\,*}$?

(4) Will there be any dependence on the initial distribution $\vec{p}(0)$?

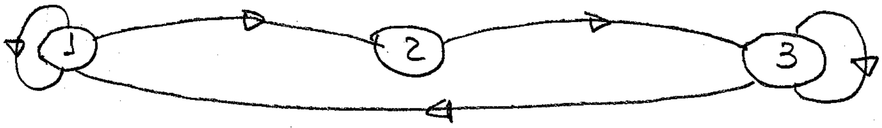
(5) How to compute $\vec{p}^{\,*}$ efficiently?

In order to answer these questions, let me first introduce some definitions.

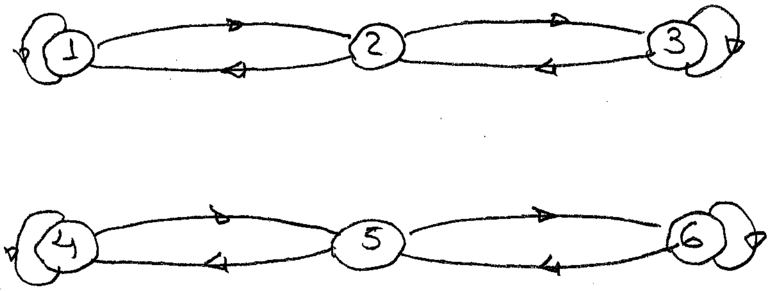
A few definitions

Consider the following examples:

(A)

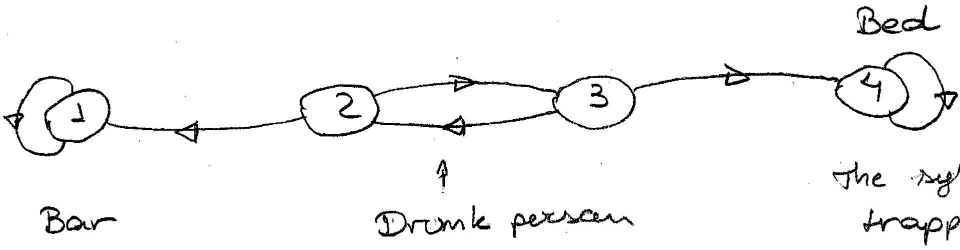


(B)



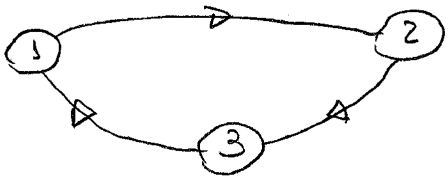
No link between (1, 2, 3) and (4, 5, 6)

(C)



The system will be trapped in 1 or 4.

(D)



This is called a cyclic chain.

→ 1 and 4 are called <u>absorbing states</u>

Irreducible chain

A chain is called irreducible if it is possible to go from anywhere to anywhere.

(A) and (D) are irreducible but (B) and (C) are not.

If you add a link $3 \to 6$ in (B), it will still not be irreducible because, if you start in the lower chain you still cannot reach the upper part.

Given a transition matrix Q, we may write the criteria of irreducibility as

"A chain is irreducible if for each (n, m) there exists an integer j such that $(Q^j)_{nm} > 0$."

This means that, in j steps it becomes possible to reach n starting from m.

Most theorems concerning the steady-state refer to irreducible matrices.

Theorems about the steady-state of irreducible chains

We have reduced the discussion to a linear algebra problem

- Steady-state : $Q\,\vec{p}^{\,*} = \vec{p}^{\,*}$ (eigenvector w/ eigenvalue 1)
- Q is a stochastic matrix ($Q_{mn} \geq 0$ and $\sum_m Q_{mn} = 1$)
- The chain is irreducible : for each (n, m) there exists an integer j such that $(Q^j)_{mn} > 0$.

The solution of this problem is the Perron - Frobenius theorem:

1) The matrix Q has only one eigenvalue equal to 1

2) All other eigenvalues satisfy $|\lambda| < 1$.

3) The eigenvector corresponding to $\lambda = 1$ is unique and all its entries are positive [they are the P_n^*].

These 3 results mean that

> " For irreducible chains the steady-state exists and is unique".

It is also possible to show the following useful property:

> $P_n^* = 1/\tau_n$
>
> τ_n = time it takes to return to state n when starting at state n.

The last question we need to answer is "will the chain always converge to $\vec{p}^*$?" Most irreducible chains will. There is only one subclass which is problematic, known as cyclic chains. An example is chain (D) in page 11. Its transition matrix will be

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{20}$$

This is a valid stochastic matrix: its entries are non-negative and the columns add to 1. It is also irreducible because you can go from anywhere to anywhere. However $Q^4 = Q$ (you may check this) so after 3 steps you are back where you started and therefore you will never reach the steady-state

To avoid these cases we consider a subclass of irreducible matrices called regular, when we compute $Q^2, Q^3, Q^4, \ldots$ the zeros in the matrix will move around. For a regular matrix, after a certain power Q^k the zeros eventually disappear. This leads us to the following result

Let Q be an irreducible stochastic matrix. If there exists an integer k for which all entries of Q^k are strictly positive, then the chain will converge to the steady-state irrespective of the initial distribution

$$\boxed{\lim_{t\to\infty} Q^t \vec{p}(0) = \vec{p}^* \quad \text{for any } \vec{p}(0)} \tag{21}$$

How to compute the steady-state using matrix multiplication
(Optional: need to know a bit of linear algebra)

According to Eq (21), if a chain is irreducible and regular (not cyclic), we may compute its steady-state by starting from an arbitrary vector $\vec{p}(0)$ and simply applying Q to it many times. Now I want to show you why this works.

Let us write the eigenvalue/eigenvector Eq for Q as

$$Q\vec{x}_i = \lambda_i \vec{x}_i \tag{22}$$

According to the Perron-Frobenius theorem, one eigenvalue will be 1 and all others will satisfy $|\lambda_i| < 1$. Choose $i=1$ as the special eigenvalue: $\lambda_1 = 1$. Then, as we discussed, $\vec{x}_1 = \vec{p}^*$.

Now suppose we start the chain with some arbitrary distribution $\vec{p}(0)$. We may always find certain coefficients c_i which allow us to expand $\vec{p}(0)$ as a linear combination of the eigenvectors $\vec{x}_i$. That is

$$\vec{p}(0) = c_1\vec{x}_1 + c_2\vec{x}_2 + c_3\vec{x}_3 + \dots \tag{23}$$

for certain numbers c_i. When we multiply $\vec{p}(0)$ by Q we will get

$$Q\vec{p}(0) = c_1 Q\vec{x}_1 + c_2 \vec{Q} x_2 + c_3 Q\vec{x}_3 + \dots$$

But using (22) we see that this will be simply

$$Q\vec{p}(0) = c_1\lambda_1\vec{x}_1 + c_2\lambda_2\vec{x}_2 + \dots \tag{24}$$

Now multiply by Q again:

$$Q^2 \vec{p}(0) = Q[Q\vec{p}(0)]$$

$$= c_1 \lambda_1 Q \vec{x}_1 + c_2 \lambda_2 Q \vec{x}_2 + \dots$$

$$= c_1 \lambda_1^2 \vec{x}_1 + c_2 \lambda_2^2 \vec{x}_2 + \dots$$

We therefore see that, if we continue like this for several iterations, we will get

$$Q^t \vec{p}(0) = c_1 \lambda_1^t \vec{x}_1 + c_2 \lambda_2^t \vec{x}_2 + \dots \qquad (25)$$

But $\lambda_1 = 1$ so $\lambda_1^t = 1$, and $|\lambda_i| < 1$ (for $i \geq 2$) which implies that λ_i^t will become smaller and smaller as we increase t. Thus, we conclude that

$$\lim_{t \to \infty} Q^t \vec{p}(0) = c_1 \vec{x}_1 \qquad (26)$$

All other terms will vanish. Since $\vec{x}_1 \propto \vec{p}^*$ we now see why $Q^t \vec{p}(0)$ will converge to the steady-state (the constant c_1 is just a normalization constant, so we choose it in a way that correctly normalizes $c_1 \vec{x}_1$).

Conclusion: if Q has only one eigenvalue 1 and all others with $|\lambda| < 1$, then the chain will converge to the steady-state. Cyclic matrices are problematic because they have other eigenvalues with $|\lambda| = 1$ so the other terms in (25) never get smaller. For instance, the eigenvalues of (20) are 1, $(-1)^{2/3}$, $(-1)^{4/3}$.

Detailed balance

The steady-state equation is $Q\vec{p}^* = \vec{p}^*$. In components this reads

$$\sum_m Q_{nm}\, p_m^* = p_n^*$$

We may write this in a more symmetric way by inserting the number 1 in the right-hand side:

$$1 = \sum_m Q_{mn}$$

We then get

$$\sum_m Q_{nm}\, p_m^* = \sum_m Q_{mn}\, p_n^* \qquad (27)$$

This is a type of global balance relation. It says that the steady-state is that state for which all transitions are balanced out.

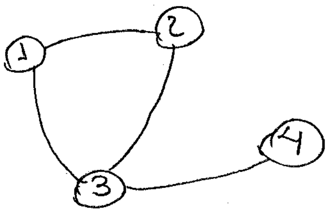
Some systems, particularly those that appear in physics, satisfy a stronger statement called detailed balance. Namely, not only the sum on the left and on the right are equal, but also its individual entries

$$\boxed{Q_{nm}\, p_m^* = Q_{mn}\, p_n^*} \qquad (28)$$

This is known as the condition of detailed balance. Not all chains satisfy it. But those related to physical processes usually do. [This is actually related to the fact that the laws of physics are usually invariant under time reversal].

Random walks on undirected graphs

Consider the following problem



I don't put arrows in the links because I always assume both transitions are possible.

The idea is: from where you are, randomly pick an edge with equal probability.

e.g.:

$$3 \to 1 \quad \text{prob } 1/3$$
$$3 \to 2 \quad \text{prob } 1/3$$
$$3 \to 3 \quad \text{prob } 1/3$$

Let n denote the different nodes ($n = 1, 2, 3, 4$) and let

d_n = degree of node n

= number of links entering node n.

For our example above $d_1 = 2$, $d_2 = 2$, $d_3 = 3$, $d_4 = 1$.

The transition probability Q_{nm} in this case is

$$\boxed{Q_{nm} = \frac{1}{d_m}} \tag{29}$$

by the above arguments. We then see that

$$\boxed{Q_{mn} d_n = Q_{nm} d_m} \tag{30}$$

Thus, the random walk on a graph satisfies detailed balance [Eq (20)].

Comparing with (20) we see that

$$P_n^* = d_n \tag{31}$$

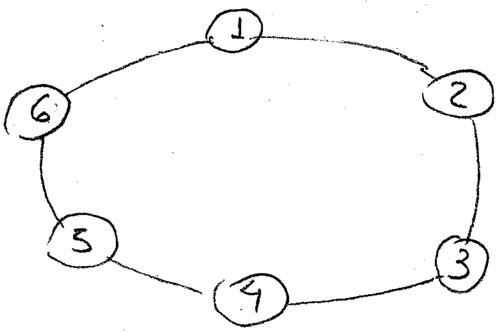
But this probability is not yet normalized, so the actual steady-state distribution will be

$$\boxed{P_n^* = \frac{d_n}{\sum_m d_m}} \tag{32}$$

This makes sense: the steady-state probability is proportional to the degree of the node. A node with many links has a higher probability of being occupied.

We may always interpret the steady-state probability as being the average fraction of the time that the system spends in that state.

Example:



At each position, the system may either move to the left or to the right with prob. 1/2

Note that this chain is not cyclic because the system can move to both sides.

The degree of each node is $d_m = 2$. Thus

$$P_m' = \frac{d_m}{\sum_m d_m} = \frac{2}{6 \times 2} = \frac{1}{6} \tag{33}$$

The system is equally likely to be found on any site. This is an example of a system with translation invariance: if you rotate the ring nothing changes. When a system has translation invariance the probabilities cannot depend on m.

Now consider what happens when we unfold the ring

The degree of the nodes are no longer all equal.

$$d_m = \begin{cases} 2 & \text{if } m = 2, \ldots, 5 \\ 1 & \text{if } m = 1 \text{ or } 6 \end{cases} \tag{34}$$

Thus $\sum_m d_m = 1 + 4 \times 2 + 1 = 10$ and the steady-state probabilities will be

$$P_m = \begin{cases} \frac{2}{10} & m = 2, 3, 4, 5 \\ \frac{1}{10} & m = 1 \text{ or } 6 \end{cases} \tag{35}$$

The prob. of finding the system at the boundaries is smaller.

The ring is what we call periodic boundary conditions (PBC) and the line is called open boundary conditions (ABC). PBC has translation invariance but ABC does not.

Google page-rank : web searches and Markov Chains

When you search for a web site, the procedure has two steps. The first is to find those sites which match your search criteria. This is easy: if you search for "Chocolate" then you only need to select all websites which have the word "chocolate" in it. The hard step is to rank these searches. To say "this website about chocolate is more important than that website".

In the beginning of the 1990s, web searching was terrible. For instance, if you created a website with the word "chocolate" written 1000 times in it, then when someone searched for "chocolate" that would be the first hit (I am not lying. I actually remember seeing exactly this).

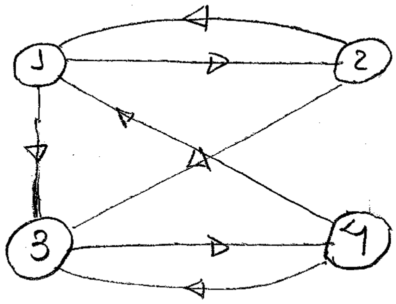
Altavista was the first engine to use the structure of the web to rank pages. It ranked a page as important when it had many links to it. But this will not be good if the link are from crappy sources.

The basic idea of Google was

Page importance = number of links to a page weighted by how important the linking pages are.

The Google PageRank algorithm was invented in 1996 by Larry Page and Sergey Brin, who were PhD students in Stanford at the time. (You can find their paper describing the algorithm online). They had the idea of modeling the web as a Markov chain on a graph. The websites were the nodes and they were linked whenever a website had in it a link to the other site.

For instance, if the world had only 4 web pages we could have



Incomming links

1 : 2

2 : 1

3 : 3

4 : 1

The idea they introduced is that of a <u>random surfer</u>, which surfs the web by randomly clicking on the links of a website with equal probability.

This generates a Markov chain on a graph. The transition matrix for our example is

$$Q = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{36}$$

The steady-state of this matrix is (you need to find the eigenvector with eigenvalue 1)

$$p^* = \begin{bmatrix} 2/9 \\ 1/9 \\ 3/9 \\ 3/9 \end{bmatrix} \tag{37}$$

Thus, sites 3 and 4 will have the highest rank, even though 3 has more links than 4.

There is, however, one problem with this: some sites don't have links to any other site: they are absorbing states (the random surfer gets trapped there).

The solution they found to this problem is to use the idea of <u>teleportation</u>. The surfer is randomly surfing the web. Each time he reaches a new site he flips a coin with probability α (a $\mathrm{Bern}(\alpha)$).

- If the coin lands 1: randomly click on any link of that site
- If the coin lands 0: move randomly to another arbitrary page in the web.

The transition matrix for this process may be written as

$$G = \alpha Q + \frac{(1-\alpha)}{M} J \tag{38}$$

where M is the total number of sites and $J = \begin{bmatrix} 1 & 1 & 1 & \cdots \\ 1 & 1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$ is a a matrix with all entries being 1 (the factor of $1/M$ is introduced to ensure that the transition matrix G continues to be a stochastic matrix).

Teleportation ensures that the chain is irreducible. To find $\vec{p}^{\,\prime}$ then all you need is to start with a random vector $\vec{p}(0)$ and multiply by G a bunch of times. As we have seen, $G^t \vec{p}(0)$ will eventually converge to $\vec{p}^{\,\prime}$.

The matrix G is enormous (there are billions of websites). But it is also incredibly <u>sparse</u> (most entries are zero). There exists sophisticated algorithms for multiplying a vector by a sparse matrix.