Stationary Distribution and Mixing Time

Josep Díaz   Maria J. Serna   Conrado Martínez
U. Politècnica de Catalunya

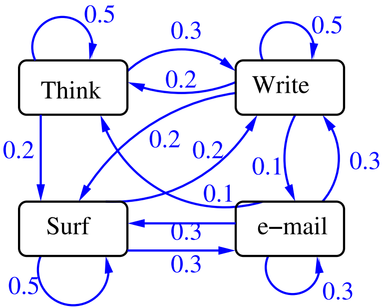
RA-MIRI 2023–2024

Stationary distribution: Writing a research paper

Recall that Markov Chains are given either by a weighted digraph, where the edge weights are the transition probabilities, or by the $|S| \times |S|$ transition probability matrix P,

Example

Writing a paper: $S = \{r, w, e, s\}$



	r	w	e	s
r	0.5	0.3	0	0.2
w	0.2	0.5	0.1	0.2
e	0.1	0.3	0.3	0.3
s	0	0.2	0.3	0.5

Stationary distributions: Writing a paper

- Suppose in the writing a paper example, the t is measured in minutes.
- For example, to see how the Markov chain will evolve after 20 minutes $\mathbb{P}[X_{20} = s | X_0 = r]$ we must compute P^{20}, and to seewhat happens 5' later $\mathbb{P}[X_{25} = s | X_{20} = s]$.
- Vectors $\pi_0 P^{20}$ and $\pi_0 P^{25}$ may be almost identical.
- This indicates that in the long run, the starting state doesn't really matter
- After a sufficiently long time t: $\pi_t \approx \pi_{t+k}$, it doesn't change when you do further steps, and this is independent of the initial distribution.
- That is, for sufficient large t, the vector distribution converges to π: $\pi_{t+1} = \pi_t P$, i.e., $\Rightarrow \pi = \pi P$.

Stationary distributions

A probability vector π is called a stationary distribution over S for P if it satisfies the stationary equations

$$\pi = \pi P.$$

If a MC has a stationary distribution π, running enough time the MC, the PMF for every X_t will be close to π.

How to find the stationary distribution

Given a finite MC with finite set of states $k = |S|$, let P be the $k \times k$ matrix of transition probabilities.

The stationary distribution $\pi = (\pi[1], \dots, \pi[k])$ over S, where $\pi_i = \pi[s_i]$ is defined by

$$(\pi[1], \dots, \pi[k]) = (\pi[1], \dots, \pi[k])P.$$

Therefore we have a system of k unknowns with k equations plus an extra equation: $\sum_{i=1}^{k} \pi[i] = 1$.

Stationary distributions: Example

Example

In the writing a paper problem, we can transform $\pi = \pi\,P$ into 5 equations to get the value of π:

$$(\pi[t], \pi[w], \pi[e], \pi[s]) = (\pi[t], \pi[w], \pi[e], \pi[s]) \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix}$$

$$\begin{aligned}
\pi[r] &= .5\pi[r] + .2\pi[w] + .1\pi[e], \\
\pi[w] &= .3\pi[r] + .5\pi[w] + .3\pi[e] + .2\pi[s], \\
\pi[e] &= .1\pi[w] + .3\pi[e] + .3\pi[s], \\
\pi[s] &= .2\pi[r] + .2\pi[w] + .3\pi[e] + .3\pi[s], \\
1 &= \pi[r] + \pi[w] + \pi[e] + \pi[s],
\end{aligned}$$

which yields, $\pi = (0.170732, 0.336043, 0.181572, 0.311653)$.

Stationary distributions

- Notice that $\pi P = \pi$ means π is a left eigenvector of P —with eigenvalue=1.
- A Markov Chain with k states and transition matrix P, it has a set of $k+1$ stationary equations with k unknowns $\{\pi[1], \ldots \pi[k]\}$, which are given by $\pi = \pi P$ together with $\sum_{u=1}^{k} \pi[u] = 1$:

$$\pi[u] = \sum_{v=1}^{k} \pi[v] P_{vu}, \ \forall 1 \leqslant v \leqslant k$$

- Linear algebra tells us that such a system either has a unique solution, or infinitely many solutions.
- We want a unique stationary distribution, so we will give conditions for MC to have a unique π.
- However, for MC with a huge number of states, it is a problem to get the stationary distribution by solving stationary equations!!

Properties of Markov chains: Recurrent

We would like to know which properties a Markov chain should have to assure the existence of a unique stationary distribution, i.e. that $\lim_{t \to \infty} P^t \to P^{(\infty)}$, for some stable matrix $P^{(\infty)}$.

A state is called recurrent if any time that we leave the state, we will return to it with probability 1.

Formally, if at time t_0 the MC is in state s, s is recurrent if the probability that $\exists t > 0$ such that $X_{t_0+t} = s$ is 1. Otherwise the state is said to be transient.

A MC is said to be recurrent if every state is recurrent.

Intuitively, transience attempts to capture how “connected” a state is to the entirety of the Markov chain. If there is a possibility of leaving the state and never returning, then the state is not very connected at all, so it is known as transient.

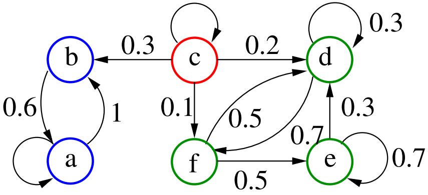
More on Recurrent and Transient MC

Alternatively, given a MC $\{X_t\}$ with state set S, a state $u \in S$ is transient if for $t > 0$,

$$\mathbb{P}[X_t = u \text{ for infinitely many } t \mid X_0 = u] = 0$$

A state $v \in S$ is recurrent if for $t > 0$,

$$\mathbb{P}[X_t = u \text{ for infinitely many } t \mid X_0 = u] = 1$$



TRANSIENT: c

RECURRENT: a,b,d,e,f

For a transient state, the number of times the chain visits s when starting at s is given by a geometric random variable in $G(p)$, where $p = \sum_{t \geqslant 1} P^t_{s,s}$.

Properties of Markov chains: Positive recurrent state

A recurrent state u has the property that the MC is expected to return to u an infinite number of times.

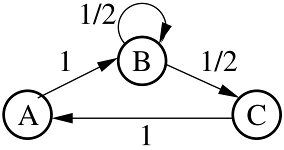
However, when restricting to finite time the MC may not return to u in a finite number of steps, which contradicts the intuition for recurrence.

We need a further finer classification of recurrence states:

If $X_t = u$ define $\tau_u = \min\{\hat{t} \mid X_{t+\hat{t}} = u\}$, as the first return time to u.

A recurrent state u is positive recurrent if $\mathbb{E}[\tau_u | X_0 = u] < \infty$.

Otherwise u is said to be a null recurrent state.



A MC with all states positive recurent.

Properties of Markov chains: Periodicity

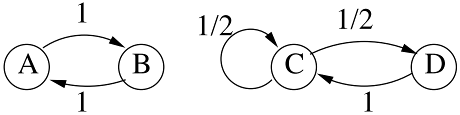
Define the period of $s_j \in S$ as $d(s_j) = \text{gcd}\{t \in \mathbb{Z}^+ \mid P^t_{s_j,s_j} > 0\}$.
So from s_j the chain can return to s_j in periods of $d(s_j)$.

Define s_j to be periodic if $d(s_j) > 1$, and s_j to be aperiodic if $d(s_j) = 1$.

A Markov chain P is periodic if every state is periodic, otherwise it is aperiodic.

Periodicity: 1st. example

A state u in a MC has period = t if only comes back to itself every t steps i.e. $P^{i}_{u,u} = 0, \forall i = t, 2t, 3t, \ldots$. Otherwise, the state is said to be aperiodic.



A,B periodic with period=2 C and D aperiodics

Notice for the left side Markov chain:

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, P^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ldots$$

$\Rightarrow \lim_{t \to \infty} P^t$ does not exist.

Periodicity: 1st. example

However, this specific Markov chain has a unique stationary distribution $\pi = (1/2, 1/2)$

Using balance eq. $(\pi[A], \pi[B]) = (\pi[A], \pi[B]) \times \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$$\begin{aligned} \pi[A] &= 0\pi[A] + 1\pi[B] \\ \pi[B] &= 1\pi[A] + 0\pi[B] \\ 1 &= \pi[A] + \pi[B] \end{aligned}$$

we get $\pi[A] = 1/2$ and $\pi[B] = 1/2$.

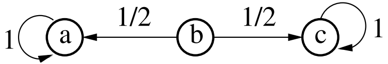
If a MC has at least one state s with self-transition $P_{s,s} > 0$ then the chain is aperiodic.

How to check if a MC is aperiodic

Given a strongly connected MC with a finite number of states,

1. If there is at least one self-transition $P_{i,i}$ in the chain, then the chain is aperiodic.
2. If you can return from i to i in t steps and in k steps, where gcd$(t, k) = 1$, then state i is aperiodic.
3. The chain is aperiodic if and only if there exists a positive integer k s.t. all entries in matrix P^k are > 0 (for all pair of states (i, j) then $P^k_{i,j} > 0$).

Properties of Markov chains: Reducibility and irreducibility



This MC is sensitive to initial state.
In this MC, $\lim_{t \to \infty} P^t$ exists, since

$$P^t = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{for all } t > 0$$

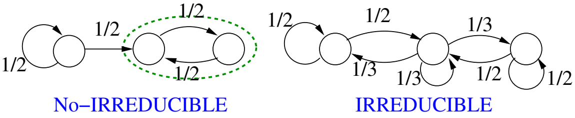
Solving the stationary equations

$$(\pi[1], \pi[2], pi[3]) = (\pi[1], \pi[2], pi[3]) \times \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix},$$

it turns out that we have infinite many stationary distributions $\pi = (p, 0, 1 - p)$.

Properties of Markov chains: Irreducibility

A finite Markov chain P is irreducible if its graph representation is strongly connected.
In an irreducible MC, the system can't be trapped in small subsets of S .



For finite Markov chains, an irreducible Markov chain is also denoted as ergodic.

Some relations among the previous classes of MC

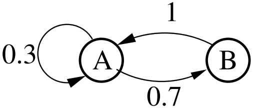
- If P is irreducible and contains a self-loop, then P is also aperiodic.
- If in a finite MC P all its states are irreducible then all the states are positive recurrent.
- If P is irreducible and finite all its states are positive recurrent, then the Markov chain has a unique stationary distribution.

Regular Markov Chain

A matrix A is defined to be regular if there is an integer $n > 0$ such that A^n contains only (strictly) positive entries.

A Markov chain is a regular Markov Chain if its transition probability matrix P is regular.

Consider the following example:



$$P = \begin{pmatrix} 0.3 & 0.7 \\ 1 & 0 \end{pmatrix} \qquad P^2 = \begin{pmatrix} 0.79 & 0.21 \\ 0.3 & 0.07 \end{pmatrix}$$

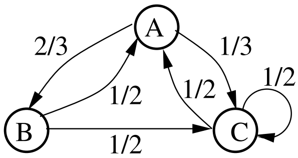
Properties of Regular MC

A finite state Markov Chain is regular if $\exists t < \infty$ such that for all states i, j, $P^t_{i,j} > 0$.

Notice that if a finite state MC is irreducible that means that for every pair of states i, j there is a t' s.t. $P^{t'}_{i,j} > 0$. If the MC is also aperiodic there is a value k s.t. for all pair of states (i, j), $P^k_{i,j} > 0$, which is exactly the definition of being regular. Therefore

Theorem

A finite state Markov chain is irreducible and aperiodic if and only if it is regular.



Markov Chains: An issue about names

- For finite state Markov chains, many people call an aperiodic, irreducible, and positive recurrent MC as ergodic, for instance Mitzenmacher & Upfal.
- However in these slides we use regular for finite MC that are aperiodic, irreducible, and positive recurrent, and reserve the name ergodic as a synonim for finite irreducible MC.
- The mathematical reason for do so is nicely explained in the link: `https://math.stackexchange.com/questions/152491/is-ergodic-markov-chain-both-irreducible-and-ape`
- N.B.: for infinite MC, regularity is not easy to define.

Fundamental Theorem of Markov Chains

Any finite, irreducible and aperiodic Markov chain P (i.e. regular) has the following properties:

1. The chain has a **unique** stationary distribution
$$\pi = (\pi[0], \pi[1], \pi[2], \ldots, \pi[n]).$$
2. $\lim_{t\to\infty} P^t$ exists and its rows are copies of the stationary distribution π.

Recall that any finite state MC has a stationary distribution, but it may not be unique.
If we have a a periodic state i, $\pi[i]$ is not necessarily the limit probability of being in state i, but the frequency of being in state i.

Markov Chain Monte Carlo technique

The Monte Carlo methods are a collection of tools for estimating values through sampling and simulations.

The Markov Chain Monte Carlo technique (MCMC) is a particular technique to sample from a desired probability distribution.

MCMC for sampling
Input: A large, but finite, set S (e.g., matching, coloring, independent set), a weight function $w : S \to \mathbb{R}^+$;
Objective: Sample $u \in S$, from a given probability distribution given by w,

$$\pi[u] \sim \frac{w(u)}{\sum_{v \in S} w(v)}$$

Technique: Construct an ad-hoc MC which converges to the distribution we want.

Why the MCMC sampling is important?

- Examining typical members of a combinatorial set (random graphs, random formulas, etc.)
- Approximate Counting: Counting the number of IS (matching, cliques, k-colorings, etc.) in a graph.
- Guessing the number of people, with a certain property, in a very large crowd.
- Combinatorial optimization, in particular heuristics.

Technique

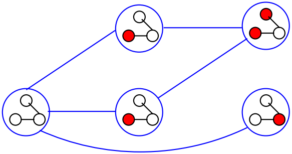
Given a state space S ($|S|$ may be very large) to form the MC, which is regular (or better symmetric):

1. Connect the state space.
2. Define carefully the transition probabilities.
3. Starting at any state u follow the MC until arriving to the stationary distribution π
   The simpler case is to aim for π be the uniform distribution.
4. Bound the maximal number steps we need to walk until arriving close enough to π.

Example: Sample the set of independent vertices in G

Given a graph $G = (V, E)$, $I \subseteq V$ is an independent set if there is no edge between any two vertices in I.

Consider the Markov chain on all the set of independent subsets of V, generated by:



We want to sample IS from the uniform distribution

Must define the appropriated transition probabilities

Example: Sampling IS in G

Given $G = (V, E)$
I_0 is an arbitrary independent set in G
To go from an independent set I_t to I_{t+1}
    choose u.a.r. $v \in V$
    if $v \in I_t$ then $I_{t+1} = I_t \backslash \{v\}$
    if $v \notin I_t$ and adding v still independent, $I_{t+1} = I_t \cup \{v\}$
    Otherwise $I_{t+1} = I_t$

Example: Sampling IS in G

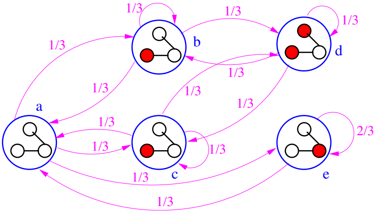
- We have a $G = (V, E)$ and $n = |V|$, An we have a set S of states, each state an independent subset of V. So $|S| \sim 2^n$.
- In the MC graph every state $I \in S$ differs from its neighbors $\mathcal{N}(I)$ in one vertex. Therefore, if $\Delta = \max\{\deg(v)|v \in I\}$ the maximum number of neighbors of any state in the MC is $\leqslant \Delta$.
- We have to define formally the transition probabilities

Transition probabilities for the MC on IS

For $I_i \in S$, with probability $1/n$ choose $v \in V$:

- If $I_i \cup \{v\}$ is not independent, stay in I_i.
- If $\{v\}$ in I_i go to new state $I_j = I_i \setminus \{v\}$.
- If $\{v\}$ is not in I_i and $I_i \cup \{v\} = I_j$ is a IS, go to I_j

Example: Sampling IS in a G



$$P = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 0 & 2/3 \end{pmatrix} \end{array}$$

$$\pi = (1/5, 1/5, 1/5, 1/5, 1/5)$$

Sampling IS in a G

Given $G = (V, E)$, $|V| = n$, and want to sample uniformly from all the N independent sets of vertices in G, including the set with 0 elements.

Make a random walk on a Markov chain on the finite but large state space $S = \{I_1, I_2, \ldots, I_N\}$, of all independent vertices in G.

Two states I_i, I_j are directly connected iff their size differs in one vertex, i.e. if their Hamming distance $|I_i \oplus I_j| = 1$.

Sampling independent vertex in a G

The transition matrix P:

$$P_{I_i,I_j} = \begin{cases} \frac{1}{n} & \text{if } |I_i \oplus I_j| = 1 \\ 1 - \frac{\mathcal{N}(I_i)}{n} & \text{if } i = j \text{ and } |I_i| \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Notice, P is aperiodic (self-loops) and irreducible (connected) so it converges to a stationary distribution.

Moreover, as $P_{I_i,I_j} = P_{I_j,I_i}$ then P is symmetric and therefore it has a uniform stationary distribution $(1/N, 1/N, 1/N, \ldots, 1/N)$.

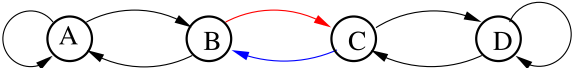
How long do we have to walk to get the stationary distribution?

Reversible Markov Chain

- For regular MC $\lim_{t\to\infty} P^t$ has all rows the same: the stationary π.
- If $|S|$ small, we can compute π by solving the stationary equations: $\pi = \pi P$.
- There is a nice property for MC, which makes more easy to compute the stationary distribution π of those MC: The reversibility.

Reversible Markov Chain

- Intuitively assume the MC below has the appropriate probabilities to have stationary distribution π. Then for sufficiently large t, $p^{t+1}_{B,C} = \pi[B]P^t_{B,C}$ (red), and $p^{t+1}_{C,B} = \pi[C]P^t_{C,B}$ (blue).
- So in stationary distribution, the rate $B \to C =$ rate $C \to B$, and this holds for every pair of adjacent states.
  i.e. For such MC, $\forall u, v \in S\; \pi[u]P_{u,v} = \pi[v]P_{v,u}$.



If a MC P has a stationary distribution π, this means π is the joint PMF for $X_0, X_1, \ldots, X_n$. Assume that we run backwards the process: if π is also the joint PMF of this time-reversal process then we say that the MC is reversible.

Reversible Markov Chain

Given a Markov Chain P, with a finite state S and a unique stationary distribution π, we say that the Markov Chain is time reversible if for all pair $u, v \in S$, it satisfies the balance equations:

$$\pi[u]P_{u,v} = \pi[v]P_{v,u}.$$

The name reversible is due to the fact that we can run the MC in the reverse and we have the same values.

The next theorem shows that if the balance equation holds for some distribution $\hat{\pi}$ then it must be a stationary distribution

Reversible Markov Chain

Theorem

Let P be Markov Chain with states S. If π is a probability vector satisfying the balance equations ($\pi[u]P_{u,v} = \pi[v]P_{v,u}$ for all $u, v \in S$) then π is a stationary distribution.

Proof

Check the stationary distribution holds, i.e. $\pi = \pi P$

$$(\pi P)[v] = \sum_{u \in S} \pi[u]P_{u,v} = \sum_{u \in S} \pi[v]P_{v,u}$$
$$= \pi[v] \sum_{u \in S} P_{v,u} = \pi[v].$$

□

Reversible Markov Chain

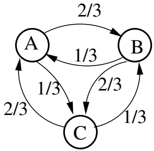
Given a finite-state MC which is reversible, to find a stationary distribution: Solve the balance equations together with the equation $\sum_{v \in S} \pi[v] = 1$.

What happens if the Markov Chain is not reversible

Not all Markov Chains are time reversible.

If there is no solution to the time reversibility equations , the way to find π is to use the stationary equations, which always yield a solution (provided the state-space in not too large).

The following MC is not reversible:



To prove it, find the stationary distribution $\pi = (1/3, 1/3, 1/3)$ and notice that $\pi[B]P_{B,C} = \frac{2}{9} \neq \frac{1}{6} = \pi[C]P_{C,B}$.

Testing if a MC is reversible: Kolmogorov's loop criterion

It is desirable to verify reversibility before finding the stationary vector π.
Recall A MC is reversible if for every finite sequence of states $i_0, i_1, i_2, \ldots i_k$ we have
$p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{k-1},i_k} p_{i_k,i_0} = p_{i_0,i_k} \cdots p_{i_1,i_0}$

Kolmogorov's loop criterion: A Markov transition matrix P is reversible iff for every loop of distinct states, the forward loop probability product equals the backward loop probability product.

But for large number of states n, the number of loops could be exponential.

Kolmogorov's loop criterion

- A two-state MC is always reversible as $p_{1,2}p_{2,1} = p_{2,1}p_{1,2}$.
- If P is symmetric (bistochastic), then. $p_{i,j} = p_{j,i}$, $\forall i, j \in S$, Kolmogorov's criterion is satisfied and P is reversible.
- If the zeros in a regular MC P are not symmetric, then the chain is not reversible
- There is a nice algorithm based in matrix operations to check if a MC P is reversible. Brill, Cheung, Hlynka, Jiang: *Reversibility checking for Markov chains*, Comm. on Stochastic Analysis, 12: 2, 129–135, (2018)

Symmetric matrix

Lemma

If P is symmetric then the MC is reversible and it has a unique stationary distribution π which is the uniform distribution, i.e. $\forall i \in S, \pi[i] = 1/n$, where $n = |S|$.

Proof

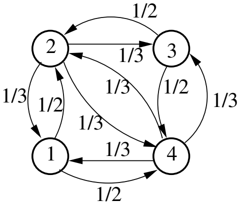
A regular MC with symmetric transition matrix is also reversible.
Then $\forall i, j \in S, \pi[i]P_{i,j} = \pi[j]P_{j,i} \Rightarrow \pi[i] = \pi[j]$. If we have n states each with the same stationary distribution then $\pi = (1/n, \dots, 1/n)$ $\square$

Example: Random walk on a graph

Given G by its adjacency matrix, a walker moves to a randomly from vertex i to a neighbor with probability $1/d(i)$.

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

Interpret random walks on G as a Markov chain and give the transition matrix P.



$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/4 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

Example: Random walk on a graph

Is $\mathbb{P}$ reversible?

Notice all non-diagonal 0s in $\mathbb{P}$ are symmetric.
There are 3 loops: (a) $(1 \to 2 \to 4 \to 1)$ (b) $(2 \to 3 \to 4 \to 2)$, (c) $(1 \to 2 \to 3 \to 4 \to 1)$.
For the (a), (b) and (c) Markov's loop criteria works, so yes.

Using the previous fact determine the stationary distribution of the MC.

As it is reversible, using the balance equations:

$$\frac{\pi[1]}{2} = \frac{\pi[2]}{3}; \frac{\pi[2]}{3} = \frac{\pi[3]}{2};$$

$$\frac{\pi[2]}{3} = \frac{\pi[4]}{3}; 1 = \pi[1] + \pi[2] + \pi[3] + \pi[4];$$

We get
$\pi = (\frac{1}{5}, \frac{3}{10}, \frac{1}{5}, \frac{3}{10})$

Number of steps: Expected first passage

Given a regular Markov chain with a set S of states , $|S| = r$, and a unique stationary distribution π,

We want to compute the expected first recurrence time for $u \in S$, $h_{u,u}$,
i.e. the expected number of steps we need so that starting from u we return for first time to u.
Intuitively, in the long run we expect the MC to be in state u a fraction of $\pi[u]$, so $h_{u,u} \sim 1/\pi[u]$

The expected first passage from u to v is denoted $h_{u,v}$. and denoted mean first passage time.

In the particular case of random walks $h_{u,v}$ is denoted as the hitting time.

Computing $h_{u,u}$ using π

Theorem

In a finite, regular Markov chain with $|S| = r$ *and a unique stationary distribution* π*, for* $u \in S$

$$h_{u,u} = \frac{1}{\pi[u]}.$$

This technique is important and it is called first step analysis: it consist in breaking down the possibilities resulting from the first step in the MC.

Proof of the Theorem

Proof

For $u, v \in S$, $h_{uv} = \mathbb{E}[\text{# steps } u \to v] = \sum_{w=1}^{k} P_{u,w}\, \mathbb{E}[\text{# steps } u \to v | 1st. \text{ step } u \to w]$

Two cases for w:

$(w = v)$ Then the expected time $u \to v$ is 1.

$(w \neq v)$

Proof of the Theorem

Proof (cont'd)

$(w \neq v)$ We take 1 step $v \to w$. By the Markovian property, we have to concentrate in state w:

$$
\begin{aligned}
h_{uv} &= P_{u,v} + \sum_{w \neq v} P_{u,w}(1 + \underbrace{\mathbb{E}[\text{time from } w \to v]}_{h_{w,v}}) \\
&= P_{u,v} + \sum_{w \neq v} P_{u,w}(1 + h_{wv}) \\
&= P_{u,v} - P_{u,v}(1 + h_{vv}) + \sum_{w=1}^{r} P_{u,w}(1 + h_{wv}) \\
&= \underbrace{-P_{u,v} h_{vv}}_{\Diamond} + \underbrace{\sum_{w=1}^{k} P_{u,w}(1 + h_{wv})}_{*} \qquad (1)
\end{aligned}
$$

Proof of the Theorem: Term $(*)$

Proof (cont'd)

Let J be the $k \times k$ matrix of 1's, then $1 + h_{wv} = (J + H)[w, v]$, where $H = (h_{v,u})$:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} h_{AA} & h_{AB} & h_{AC} \\ h_{BA} & h_{BB} & h_{BC} \\ h_{CA} & h_{CB} & h_{CC} \end{pmatrix} = \begin{pmatrix} 1 + h_{AA} & 1 + h_{AB} & 1 + h \\ 1 + h_{BA} & 1 + h_{BB} & 1 + h \\ 1 + h_{CA} & 1 + h_{CB} & 1 + h \end{pmatrix}$$

$$\sum_{w=1}^{k} h_{u,w}(1+h_{wv}) = \sum_{w=1}^{k} H[u, w](J+H)[w, v] = (H\times(J+H))[u, v].$$

So the sum is just the entry (w, v) in the matrix $H \times (J + H)$.

Proof of the Theorem: Term (◊)

Proof (cont'd)

Introduce $r \times r$ diagonal matrix D, where $D_{v,v} = h_{v,v}$,
so $P_{u,v} h_{v,v} = (P \times D)[u, v]$.

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix} \times \begin{pmatrix} h_{AA} & 0 & 0 \\ 0 & h_{BB} & 0 \\ 0 & 0 & h_{CC} \end{pmatrix} = \begin{pmatrix} h_{AA}/3 & h_{BB}/3 & h_{CC}/3 \\ h_{AA}/2 & 0 & h_{CC}/2 \\ h_{AA}/2 & h_{BB}/2 & 0 \end{pmatrix}$$

Ending the proof

Proof (cont'd)

Substituting $\Diamond$ and $*$ in equation (1): $h_{u,v} = -(P \times D)[u,v] + (H \times (J+H))[u,v]$.

As it is true $\forall(u,v)$, we have $H = -PD + P(J+H) = -PD + PJ + PH$

Multiply both sides by the stationary distribution: $\pi H = -\pi PD + \pi PJ + \pi PH$.
But by the stationary equation: $\pi P = \pi \Rightarrow \pi H = -\pi D + \pi J + \pi H \Rightarrow \pi D = \pi J$

Notice πJ is just the k-dimensional vector $(1, 1, \ldots, 1)$ $(\sum_i \pi[i] = 1)$

$\pi D = (\pi[1]h_{11}, \pi[2]h_{22}, \ldots, \pi[k]h_{kk}) \Rightarrow \pi[u] = \frac{1}{h_{u,u}}$ $\square$