Markov Chains: stationary distribution

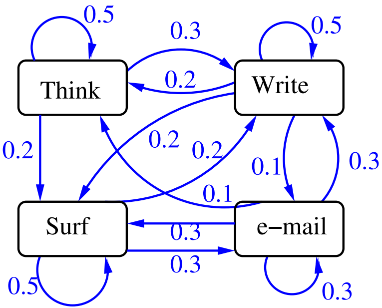
RA-MIRI

QT Curs 2020-2021

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 $S = \{r, w, e, s\}$



$$\begin{array}{c c} & \begin{array}{cccc} r & w & e & s \end{array} \\ \begin{array}{c} r \\ w \\ e \\ s \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0 & 0.2 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.3 & 0.3 \\ 0 & 0.2 & 0.3 & 0.5 \end{pmatrix} \end{array}$$

Stationary distributions: Writing a paper

- Suppose in the writing a paper example, the t is measured in minutes.
- To see how the Markov chain will evolve after 20 minutes i.e. $\mathbf{Pr}\left[X_{20} = s | X_0 = r\right]$ we must compute P^{20}, and to see if 5' later $\mathbf{Pr}\left[X_{25} = s | X_{20} = s\right]$.
- 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,
- which implies that after a sufficiently long t: $\pi_t = \pi_{t+k}$, 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 a 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 $\pi = (\pi[1], \ldots, \pi[k])$ over S, where $\pi_i = \pi[s_i]$. is defined by

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

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

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 $\{\pi[1], \dots \pi[n]\}$ 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 $\{\pi[1], \dots \pi[n]\}$, which are given by $\pi = \pi P$ together with $\sum_{u=1}^{k} \pi[u] = 1$:

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

- 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 that 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 $\lim_{t\to\infty} P_t \to$ *a stable matrix.*

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

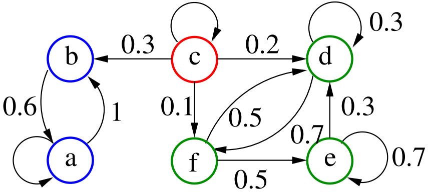
Formally, if at time t_0 the MC is in state s, s is recurrent if $\exists t$ such that $P_{s,s}^{t_0+t} = 1$. Otherwise the state is said to be transient.

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 $u \in S$ is transient if for $t > 0$,
$\mathbf{Pr}[X_t = u \text{ for infinitely many } t \mid X_0 = u] = 0.$
A $v \in S$ is recurrent if for $t > 0$,
$\mathbf{Pr}[X_t = u \text{ for infinitely many } t \mid X_0 = u] = 1.$



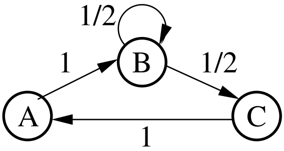
TRANSIENT: c

RECURRENT: a,b,d,e,f

For 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 $p = \sum_{t \geq 1} P^t_{s,s}$.

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.
Define a recurrent state u to be positive recurrent if $\mathbf{E}[\tau_u | X_0 = u] < \infty$. Otherwise u is said to be null recurrent state.



A MC with all states positive recurent.

Properties of Markov chains: Periodicity

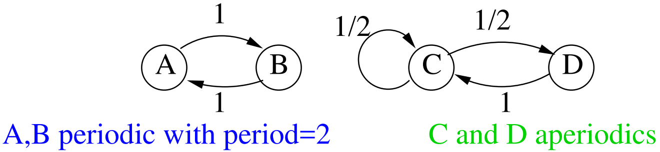
Define the period of $s_j \in S$ as $d(s_j) = \gcd\{t \in \mathbb{Z}^+ \mid P^n_{s_j,s_j} > 0\}$.
So from s_j the chain can return to s_j in periods of $d(s_j)$.

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.



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 balence 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 $\pi[B] = 1/2$.

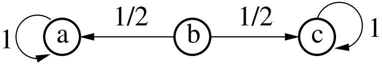
If one MC has at least one state s with self-transition $P_{s,s} > 0$ then the chain is aperiodic.

How to check if a MC is aperiodic

Given an irreducible 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, $\forall t$, $\lim_{t\to\infty} P^t$ exists,

$$P^t = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix},$$

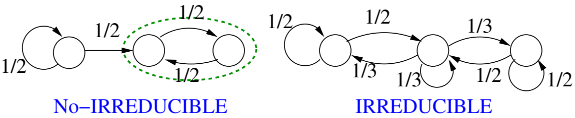
Solving the stationary equations

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

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 W is strongly connected.
In irreducible W, 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

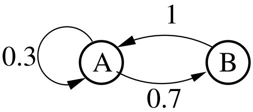
- If P is irreducible and contains a self-loop, then P is also aperiodic.
- If in a finite MC P all its states are irreducible then all the states are positive recurrent.
- If P is irreducible and finite all its states are positive recurrent, then the Markov chain has a unique stationary distribution.

Regular Markov Chain

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

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

Consider the following example:



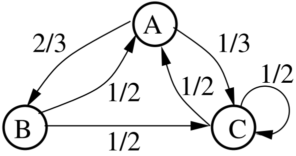
$$P = \begin{pmatrix} 0.3 & 0.7 \\ 1 & 0 \end{pmatrix} \qquad P^2 = \begin{pmatrix} 0.79 & 0.21 \\ 0.3 & 0.07 \end{pmatrix}$$

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 denotes a that is aperiodic, irreducible, and positive recurrent as ergodic, as for instance in MU.
- However in this slides we use regular for finite MC that are aperiodic, irreducible, and positive recurrent, and reserve the name ergodic for 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-aperiodic
- However 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 row 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 (matching, coloring, independent sets), 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 and 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

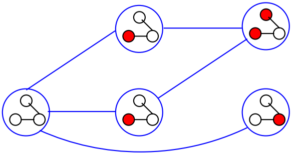
Given a state space S ($|S|$ may be very large) to form the MC, which is regular (or better symmetric):

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 streps we need to walk until arriving to π.

Example: Sample the set of independent vertices in G

Given a graph $G = (V, E)$ the $I \subseteq V$ is 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 state, each state an independent subset of V. So $|S| \sim 2^n$.
- In the MC graph every state $I \in S$ of differs from its neighbors $\mathcal{N}(I)$ in one vertex. Therefore, if $\Delta = \max\{d(I)|I \in S\}$ the maxim number of neighbors of any state in the MC is $\leq \Delta$.
- To define the transition probabilities:
  implementing directly a RW in the graphs of the MC may not work as not all states have the same degree.

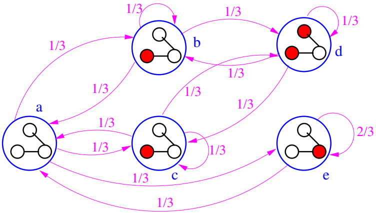
Transition probabilities for the MC on IS

Take for any $I_i, I_j \in S$, $P_{I_i,I_j} = 1/\Delta = 1/n$.

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 without v.
- If $\{v\}$ is not in I_i and form an i.s. adding v, i.e. go to I_j.

Example: Sampling IS in a G



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

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

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 $S = \{I_1, I_2, \dots, I_N\}$, 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:

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

Notice, P is aperiodic (self-loops) and irreversible (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 go in the RW to get the stationary distribution?