Chapter 8. RANDOM WALKS - MARKOV CHAINS

Random walks on graphs are a very simple, interesting and fundamental tool with surprisingly many applications in mathematics, informatics and also in natural sciences. Design and analysis of randomized algorithms is one of them.

The concept of a random walk is closely related with that of Markov chain – one of the key concepts of discrete stochastic processes.

In physical sciences, Markov chains provide a fundamental model for the emergence of global properties from local interactions.

In informatics, random walks provide a general paradigm for random exploration of an exponentially large combinatorial structures (for example graphs), by a sequence of simple and local transitions.

Developments of Markov chains

The concept of Markov chain introduced Andreei Andreevich Markov in 1906 to consider the case of having a sequence of random experiments in which result of any experiment depends also on the result of the previous experiment.

Before that only such sequences of random experiments were considered where the result of each random experiment was fully independent from all previous experiments.

UNIVERSALITY of QUANTUM RANDOM WALKS

It can be shown that any

quantum evolution

can be seen as so-called

continuous quantum walk.

BASIC CONCEPTS I

Notation Let $G = (V, E)$ be a connected, undirected, graph with n nodes (in V) and m edges (in E). For a node $v \in V$, let $\Gamma_G(v)$ denote the set of neighbors of v in G.

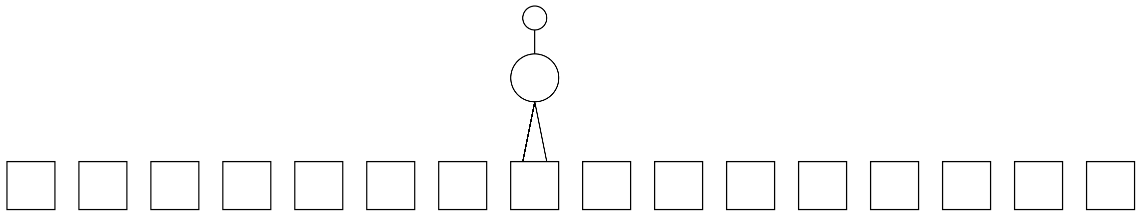
A **random walk** on G is the following sequence of moves of the process that starts in a node v_0:

> a neighbor v_1 of v_0 is chosen, randomly and independently, and then the process moves (walks) from v_0 to v_1; afterwards a neighbor v_2 of v_1 is chosen randomly and independently and the process walks from v_1 to v_2; the process continues to walk this way $\ldots$for ever.

Typical problems to investigate for a given graph G:

- What is the expected number of steps to get from a given node u to a given node v.
- What is the expected number of steps needed to visit all nodes of G when starting in a given node u.

DRUNKARD's WALK



Let a drunken seaman walk on a linear, both sides infinite graph, each time choosing a step right or left with the same probability.

What are probabilities for such a drunken man to be in a particular position after some stepsi in case he starts in some other position.

EXAMPLE

Let

- $G = K_n$ be the complete graph of n nodes
- $u \neq v$ be any two vertices of G.

It holds:

- The expected number of steps of a random walk that begins in u and ends when first reaching v is
$$n - 1.$$
- The expected number of steps to visit all nodes in G starting from the node u is
$$(n-1)H_n,$$
where H_n is Harmonic number

A RELATED PROBLEM - COUPON SELECTION

Coupon selector problem: There are n types of coupons and at each time a coupon is chosen randomly and returned.

It has been shown that the average number of trials needed to have a coupon of each type is

$$nH_n.$$

EXAMPLE

Let us consider graph K_3 with the initial probabilities of its three nodes, say A0, A1, A2, being

$$p_0, p_1, p_2$$

and let for each node and its edge the probability of the transmission through this edge be $1/2$.

If p is the initial probability of a node X, then probability of being in the same node after one step is

$$P_1 = (1-p)\frac{1}{2} = \frac{1}{2} - \frac{p}{2},$$

after two steps is

$$P_2 = (1-(1-p)\frac{1}{2})\frac{1}{2} = \frac{1}{2} - \frac{1}{4} + \frac{p}{4}$$

and after i steps the probability is

$$P_i = \sum_{j=1}^{i} (-1)^{j+1}\frac{1}{2^j} + (-1)^i \frac{p}{2^i}$$

and therefore

$$\lim_{i\to\infty} P_i = \frac{1}{3}$$

EXAMPLE – 2-SATISFIABILITY

A simple polynomial-time (Monte Carlo) algorithm will be given to solve 2-SAT problem (for a formula in conjunctive normal form with each clause having two literals) satisfiable with a single assignment.

Algorithm

1. Start with an arbitrary assignment.
2. **while** there is an unsatisfied clause C, choose randomly one of two literals of C and complement its value in the current assignment.

Theorem The expected number of steps of the above algorithm at finding a satisfying assignment is $\mathcal{O}(n^2)$ (where n is the number of variables).

Relation to a random walk on the line

Let A be a particular satisfying assignment.

The progress of the above algorithm can be represented by a particle moving between integers $\{0, 1, \ldots, n\}$ on the real line. The position of the particle will always indicate how many variables in the current assignment have the correct value (as in A).

Crucial observation. In an unsatisfied clause at least one of two literals has an incorrect value. Therefore at each step of the algorithm with probability $\frac{1}{2}$ we increase by one the number of variables having their correct value; with probability $\frac{1}{2}$ the number of variables having correct value is decreased by one. The motion of the particle therefore resembles a random walk on the line (that is on the linear graph).

MARKOV CHAINS - 1st DEFINITION

A **Markov chain** is a discrete-time stochastic process defined over a set of states S in terms of a matrix P of transition probabilities

$$P(i,j) \; = \; p_{ij} = \text{ the probability that the next state will be } j \text{ if the current state is } i.$$

Probability conditions: For any i, j it has to hold

$$0 \leq p_{ij} \leq 1 \text{ and } \sum_j p_{ij} = 1.$$

Denote X_t the state of the Markov chain at time t.

The sequence $\{X_t\}_{t=0}^{\infty}$, specifying the history of the evolution of the Markov chain at time t, has the following **memoryless property**:

The future behaviors of a Markov chain depends on its current state, and not how the chain arrived at the current state. That is, it holds

$$Pr[X_{t+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t = i] = Pr[X_{t+1} = j | X_t = i] = p_{ij}.$$

Note: Markov chains do not need to have prespecified initial states.

In general, initial states are chosen according to some probability distribution X_0 over S. Such a distribution is called the **initial (probability) distribution**.

MARKOV CHAIN –2nd DEFINITION

A Markov chain is a discrete time stochastic process $\{X_k\}_{k\geq 0}$ of random variables with values in a countable set I such that for every $k \geq 1$ and every $i_0, i_1, \ldots, i_k$ from I we have

$$Pr[X_k = i_k \,|\, X_{k-1} = i_{k-1}, X_{k-2} = i_{k-2}, \ldots, X_0 = i_0] = Pr[X_k = i_k \,|\, X_{k-1} = i_{k-1}] = p_{i_{k-1}i_k}.$$

The matrix $P(i,j) = p_{ij}$ is called the **transition matrix**.

k-**steps transition probabilities** $p_{ij}^{(k)}$ are defined by

$$p_{ij}^{(k)} = Pr[X_{m+k} = j \,|\, X_m = i]$$

and they do not depend on m.

If we define a matrix $P^{(k)}$ by $P^{(k)}(i,j) = p_{ij}^{(k)}$, then

$$P^{(k+m)} = P^{(k)}P^{(m)}.$$

The matrix $P^{(k)}$ is said to be the k-steps transition matrix.

Basic **reachability concepts**:

- A state j is called **reachable (accessible)** from the state i if there is a $k > 0$ such that $p_{ij}^{(k)} > 0$.
- We say that states i and j are mutually reachable if i is reachable from j and vice verse.
- Markov chain is called **irreducible**, if any two of its states are mutually reachable.

CREATURE EATING HABITS

Systems represented by Markov chains change randomly and therefore it is generally impossible to determine the exact state of the system in the future. However, we often can determine various useful statistical properties of such systems.

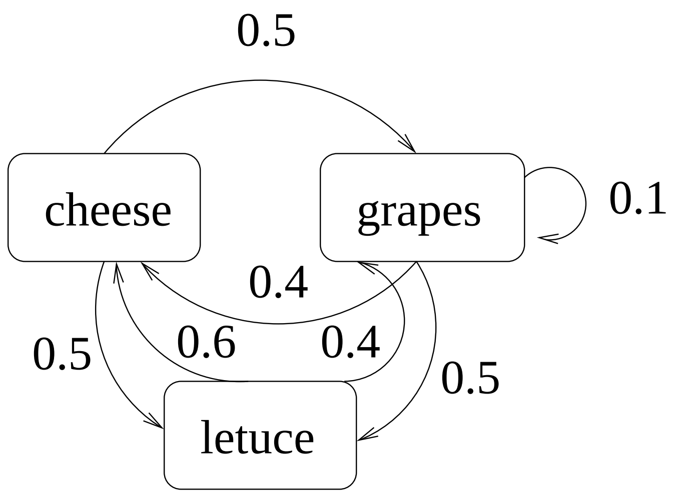
Example Let a creature eat once a day, either grapes, or cheese, or lettuce, according to the following rules:

- If it ate cheese yesterday it will not eat it today and will eat lettuce and grapes with the same probability 0.5.
- If it ate grapes yesterday, it will eat today grapes with probability 0.1 cheese with probability 0.4 and lettuce with probability 0.5
- If it ate lettuce yesterday, it will eat grapes with probability 0.4 and cheese with probability 0.6.

These eating habits can be modelled by Markov chain.One statistical property that can be computed is the percentage of days it eats grapes (or cheese).

MARKOV CHAIN for EATING HABITS

Markov chains are often described by a directed graph with edges labelled by probabilities for going from one state to another one.



Eating habits of a creature.

Ehrenfest model

There are two urns that, between them, contain four balls. At each step, one of the balls is chosen at random and moved to other urn.

If we choose as states numer of balls in the first urn then transition matrix, where rows and columns are labelled (from top to bottom and from left to right) $\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}$, $\mathbf{3}$, $\mathbf{4}$ as follows

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

ABSORBING DRUNKARD's WALK

A drunk man walks on a 5 nodes $(0, 1, 2, 3, 4)$ linear graph, where leftmost node (0) is his home and rightmost one (4) is a bar. If he is at home or in bar he keeps staying there. Otherwise he moves with probability 1/2 to left node and probability 1/2 to right node.

The transition matrix has therefore the form

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

where raws and columns are labeled by $0, 1, 2, 3, 4$

ABSORBING MARKOV CHAINS

Definition A state s_i of a Markov chain is called absorbing if it is impossible to leave it (i.e. $p_{ii} = 1$). A Markov chain is called absorbing if it has at least one absorbing state, and if from any state one can go to an absorbing state, in some number of steps. A state that is not absorbing is called transient.

The Drunkard's walk is an example of an absorbing Markov chain.

If the transition matrix of a Markov walk has a absorbing states and t transient states, one can renumber states so that all first t rows and columns represent transient state and remaing one absorbing states. The matrix has then the following form:

$$P = \begin{pmatrix} Q & R \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where Q is a $t \times t$ matrix, R is a $t \times a$ matrix, $\mathbf{O}$ is a $a \times t$ zero matrix and I is $a \times a$ identity matrix, with first t rows and columns labeled by transient states.
This means

$$P^n = \begin{pmatrix} Q^n & * \\ \mathbf{O} & \mathbf{I} \end{pmatrix}$$

where $*$ is a matrix the precise form of which will not be important in the following.

Properties of the transition matrix of absorbing Markov chains.

Theorem. If P is the transition matrix of an absorbing Markov chain in the normal form shown on previous page. Then it holds

- $\lim_{n\to\infty} Q^n = \mathbf{O}$, that is the rpobability that the process will be absorbed is 1.
- The inverse matrix to the matrix $I - Q$ is the matrix
$$N = I + Q + Q^2 + \ldots$$
- Let t_i be the expected number of steps before the chain is absorbed, given that the chain starts in the state S_i and let **t** be the column vector whose ith entry is t_i. Then
$$t = Nc$$
where c is a column vector all of whose entries are 1.
- Let b_{ij} be the probability that an absorbing chain will be absorbed in the absorbing state s_j if it starts in a transient state s_i. Let B be the matrix with entries b_{ij}. Then B is an $t \times a$ matrix
$$B = NR,$$
where R is as in the canonical form.

BASIC CONCEPTS – II

Given an initial state $X_0 = i$, the **probability that the first transition into state j occurs at time t** is given by

$$r_{ij}^{(t)} = Pr[X_t = j \text{ and } X_s \neq j \text{ for } 1 \leq s < t \mid X_0 = i].$$

Given an initial state $X_0 = i$, then the **probability that there is a visit to state j at some time $t > 0$** is given by

$$f_{ij} = \sum_{t>0} r_{ij}^{(t)}.$$

f_{ii} is the probability that Markov chain will return to the state i at least once when started in the state i.

If $f_{ii} = 1$ the state i is called **recurrent (vracajuci sa)**; otherwise it is called **transient (prechodný)**. A markov chain is recurrent if every its state is recurrent.

All states of an irreducible Markov chain are recurrent.

If $f_{ii} < 1$, then each time the chain is in the state i, with probability $1 - f_{ii}$ it will never return again to i. It therefore holds:

$$Pr[\text{The number of visits to } i \text{ from } i \text{ equals } k] = f_{ii}^k(1 - f_{ii}).$$

BASIC CONCEPTS III

Denote as the hitting time (čas dosiahnutia/riešenia) h_{ij} the expected number of steps needed to visit the state/node j for the first time when starting from the state/node i. Clearly, it holds

$$h_{ij} = \sum_{t>0} t r_{ij}^{(t)}.$$

If $h_{ii} < \infty$, then the state i is called positive (non-null) recurrent; otherwise it is called null-recurrent.

If a state i is reachable from itself, then the greatest common divisor of the set of positive k's such that $p_{ii}^{(k)} > 0$, is called the period of i and is denoted by d_i. If $d_i = 1$, then the state i is said to be aperiodic.

A finite Markov chain all states of which are aperiodic and positive recurrent is called ergodic.

EQUIVALENT DEFINITIONS

A state i has period k if any return to the state i must occur in multiples of k time steps.

If $k = 1$, then the state is called aperiodic - returns to state i can occur at irregular times.

A state i is said to be transient, if given that we start in the state i, there is non-zero probability that we will never return to i.

If a state i is not transient, then it is called recurrent.

EXAMPLE of a Markov chain with null-recurrent states

Consider a Markov chain whose states are all positive integers.

From each state i the next states are the state $i+1$ (with probability $\frac{i}{i+1}$) and the state 1 (with probability $\frac{1}{i+1}$)

Starting at state 1, the probability of not having returned to state 1 within the first t steps is

$$\prod_{j=1}^{t} \frac{j}{j+1} = \frac{1}{t+1}.$$

Hence the probability of never returning to state 1 from 1 is 0, and state 1 is recurrent. It holds also

$$r_{1,1}^{t} = \frac{1}{t} \cdot \frac{1}{t+1} = \frac{1}{t(t+1)}.$$

However, the expected number of steps until the first return to state 1 from state 1 is

$$h_{1,1} = \sum_{t=1}^{\infty} t \cdot r_{1,1}^{t} = \sum_{t=1}^{\infty} \frac{1}{t+1}.$$

which is unbounded.

However, it holds: In a finite Markov chain at least one state is recurrent and all recurrent states are positive recurrent.

Ergodic Markov Chains

An equivalent definition of ergodic Markov chains.

Definition A Markov chain with a transition matrix P is called ergodic if it is possible to go from every state to every state and there is an integer n such that all entries of the matrix P^n are positive.

THE ERGODIC THEOREM

Suppose $\{X_k \mid k \geq 0\}$ is an ergodic Markov chain with the set of states $S = \{1, \ldots, n\}$. Then, it holds:

- There exists a vector $\pi = (\pi_1, \ldots, \pi_n)$ such that for every $i, j \in I$ it holds

  $$\pi_j = \lim_{k \to \infty} p_{ij}^{(k)}$$

  and the limit does not depend on i.

- The vector π is the only non-negative solution of the system of linear equalities

  $$\pi_j = \sum_{i=1}^{n} \pi_i p_{ij}, \quad \sum_{j=1}^{n} \pi_j = 1.$$

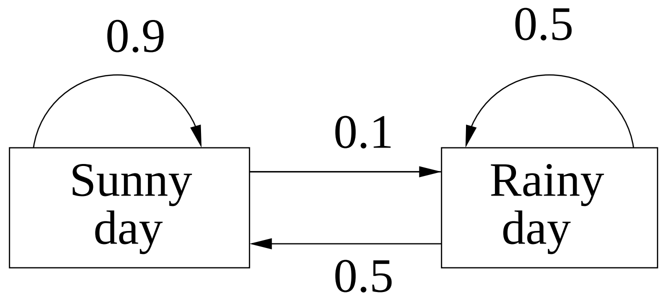
- The vector π satisfies the identity $\pi = \pi P$.
- For every $1 \leq i \leq n$, it holds that $f_{ii} = 1$ and $h_{ii} = \frac{1}{\pi_i}$.
- If $N(i,t)$ denotes the number of visits to the state i within the t first steps, then

  $$\lim_{t \to \infty} \frac{N(i,t)}{t} = \pi_i.$$

Implications: Ergodic Markov chains always “forget”, after a while, their initial probability distribution.

WEATHER FORECAST

The following Markov chain depicts probabilities for going from sunny days to rainy and vice verse.



Transition matrix has the form

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$$

and the stationary probability distribution is $\pi = (0.833, 0.167)$. It holds $\pi P = \pi$.

EXAMPLE - QUEUE

Let us consider bounded queue in which each time moment exactly one of the following steps happen.

- If a queue has fewer than n customers, then with probability λ a new customer join the queue.
- If the queue is not empty, then with probability μ the head of the line is served and leaves the queue.
- With remaining probability, the queue is unchanged.

Let us consider stochastic process with X_t being the number of customers at time t. This is a Markov chain and its transition matrix has the following non-zero entries:

$$
\begin{aligned}
P_{i,i+1} &= \lambda \text{ if } i < n \\
P_{i,i-1} &= \mu \text{ if } i > 0 \\
P_{i,i} &= \begin{cases} 1-\lambda & \text{if } i = 0 \\ 1-\lambda-\mu & \text{if } 1 \le i \le n-1 \\ 1-\mu & \text{if } i = n \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
P_{i,i+1} &= \lambda \text{ if } i < n \\
P_{i,i-1} &= \mu \text{ if } i > 0 \\
P_{i,i} &= \begin{cases} 1-\lambda & \text{if } i = 0 \\ 1-\lambda-\mu & \text{if } 1 \le i \le n-1 \\ 1-\mu & \text{if } i = n \end{cases}
\end{aligned}
$$

This Markov chain is ergodic and therefore it has a unique stationary distribution π. It holds

$$
\begin{aligned}
\pi_0 &= \pi_0(1-\lambda) + \pi_1\mu \\
\pi_i &= \pi_{i-1}\lambda + \pi_i(1-\lambda\mu) + \pi_{i+1}\mu, 1 \le i \le n-1 \\
\pi_n &= \pi_{n-1}\lambda + \pi_n(1-\mu)
\end{aligned}
$$

From that one can show that

$$\pi_i = \pi_0 \left(\frac{\lambda}{\mu}\right)^i$$

is a solution of the above system of equations. Since

$$1 = \sum_{i=0}^{n} \pi_i = \sum_{i=0}^{n} \pi_0 \left(\frac{\lambda}{\mu}\right)^i$$

it holds

$$\pi_0 = \frac{1}{\sum_{i=0}^{n} (\lambda/\mu)^i}$$

MARKOV CHAINS for GRAPHS

Let $G = (V, E)$ be a connected, non-bipartite, and undirected graph with $|V| = n$ and $|E| = m$. G induces a Markov chain, denoted by M_G, states of which are vertices of G and for any two vertices $u, v \in V$

$$P(u,v) = \begin{cases} \frac{1}{d(u)}, & \text{if } (u,v) \in E; \\ 0, & \text{otherwise;} \end{cases}$$

where $d(u)$ is the degree of u (in G).

Properties of M_G

- M_G is irreducible.
- Periodicity of M_G is the greatest common divisor of the length of all closed walks in G.
- Since G is undirected, there are closed walks of length 2;
- Since G is non-bipartite, it has odd cycles and therefore the *greatest common divisor* of all closed walks is 1. Hence, M_G **is aperiodic**.

STATIONARY DISTRIBUTIONS for MARKOV CHAINS on GRAPHS

Ergodic theorem therefore implies that M_G has a unique stationary distribution π. This π is easy to determine. Indeed, it holds

Lemma For all $v \in V$, $\pi_v = \frac{d(v)}{2m}$.
Proof Let $[\pi P]_v$ be the v-th component of πP. Then

$$\pi_v = [\pi P]_v = \sum_u \pi_u P(u,v) = \sum_{(u,v)\in E} \frac{d(u)}{2m} \times \frac{1}{d(u)} = \sum_{(u,v)\in E} \frac{1}{2m} = \frac{d(v)}{2m}.$$

Corollary For all $v \in V$, $h_{vv} = \frac{1}{\pi_v} = \frac{2m}{d(v)}$.
Comment Google's page ranking algorithm is essentially a Markov chain over the graph of the web.

TIME REVERSIBLE MARKOV CHAINS

Ergodic Markov chains with the stationary distribution π and transition probabilities $P_{i,j}$ is called time reversible, if for any states i, j it holds

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

For such chains stationary distribution is easy to compute by means of the following theorem.

Theorem Consider ergodic Markov chain with states $\{1, \ldots, n\}$ and the transition matrix P. If there are non-negative numbers $\bar{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ such that $\sum \pi_i = 1$ and for any pair of states i, j it holds

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

then π is the stationary distribution corresponding to P.

Proof Consider the j-th entry of $\bar{\pi}$. Conditions of theorem imply for any j

$$\sum_{i=1}^{n} \pi_i P_{i,j} = \sum_{i=1}^{n} \pi_j P_{i,j} = \pi_j$$

and therefore it holds $\pi P = \pi$ and therefore π is the stationary distribution.

MONTE CARLO ESTIMATION OF π

- **Let $Z = (X, Y)$ be a point chosen randomly in a 2×2 square centered in $(0, 0)$.**
- **This is equivalent to choosing X and Y randomly from interval $[-1, 1]$.**
- **Let Z be considered as random variable that has value 1 (0) if the point (X, Y) lies in the circle of radius 1 centered in the point $(0, 0)$.**
- **Clearly**

$$\boldsymbol{Pr}(Z = 1) = \frac{\pi}{4}$$

- **If we perform such an experiment m times and Z_i be the value of Z at the ith run, and $W = \sum_{i=1}^{m} Z_i$, then**

$$\mathbf{E}[W] = \mathbf{E}\left[\sum_{i=1}^{m} Z_i\right] = \sum_{i=1}^{m} \mathbf{E}[Z_i] = \frac{m\pi}{4}$$

and therefore $W' = (4/m)W$ is a natural estimation for π

- **How good is this estimation? An application of the first Chernoff bound from the page 5 of Chapter 6 gives**

$$\begin{aligned}\boldsymbol{Pr}(|W' - \pi| \geq \varepsilon\pi) &= \boldsymbol{Pr}\left(\left|W - \frac{m\pi}{4}\right| \geq \frac{\varepsilon m\pi}{4}\right) \\ &= \boldsymbol{Pr}([W - \mathbf{E}[W] \geq \varepsilon\mathbf{E}[W]) \\ &\leq 2e^{-m\pi\varepsilon^2/12}\end{aligned}$$

- **Therefore, by taking m sufficiently large we can get arbitrarily good approximation of π**

FROM SAMPLING to COUNTING

In the previous example the Monte Carlo method used a uniform sampling in a square to achieve a counting - to determine the value of π.

This method can be generalised. In many cases we can count the number of sets with a special property P in a graph in case we can generate almost uniform sample from the sets with the property P.

In various other cases we can do efficient computation provided we can implement sampling having a fixed desired probability distribution.

Markov chain Monte Carlo method provides a very general approach to sample from a desired probability distribution.

The basic idea is to create an Ergodic Markov Chain whose set of states is the sample space and whose stationary distribution is the required sampling distribution.

Let $X_0, X_1, \ldots$ be a run of such a chain. The markov chain converges to the stationary distribution from any state X_0 and so after a sufficiently large number of steps r, the distribution of the state X_r will be close to the stationary distribution and so it can be used as a sample. We can repeat the same argument starting with X_r and getting to X_{2r} and so on.

DESIGN of MARKOV CHAINS with UNIFORM STAT. DISTRIB.

We show how to construct a Markov Chain with a stationary distribution that is uniform over the given state space Ω.

The first step is to define on Ω a neighbourhood relation - that is to make out of Ω a graph - and in such a way that graph is irreducible.

The second step is to define transition probabilities in such a way that the resulting stationary distribution is uniform.

The next lemma show that this can be done in case we introduce also self-loops.

Notation For any $x \in \Omega$ let $N(x)$ is the set of neighbours in the created graph and let $M > N = \max_{x\in\Omega} |N(x)|$.

Lemma For a finite state space Ω and given neighbourhood structure and any $M > N$ design a Markov chain such that for any $x, y \in \Omega$

$$P_{x,y} = \begin{cases} 1/M & \textbf{if } x \neq y, y \in N(x) \\ 0 & \textbf{if } x \neq y, y \notin N(x) \\ 1 - N(x)/M & \textbf{if } x = y \end{cases}$$

Then if this chain is irreducible and aperiodic, then the stationary distribution is the uniform distribution.

It is easy to see that the chain is time reversible and so we can apply corresponding theorem from slide 29. Therefore for any x, $\pi_x = \frac{1}{|\Omega|}$.

EXAMPLE - INDEPENDENT SETS of a GRAPH

Consider the Markov chain, whose states are independent sets in a graph $G = (V, E)$.

1. Let X_0 be an arbitrary independent set in G.
2. To compute X_{i+1} do the following
   - choose a node v uniformly and randomly from V;
   - if $v \in X_i$, then $X_{i+1} = X_i - \{v\}$;
   - if $v \notin X_i$ and if adding v to X_i still gives an independent set, then $X_{i+1} = X_i \cup \{v\}$;
   - otherwise set $X_{i+1} = X_i$

Using the construction from previous lemma one can show that if x, y are neighbouring independent sets then $P_{x,y} = 1/|V|$ and the stationary distribution is the uniform distribution.

EXERCISES

- **From a bag of white and black balls you should pick up a white ball with probability $\frac{1}{3}$ and a black ball with probability $\frac{2}{3}$. How can you do that?**
- **You have 11 boxes of balls. You should pick**
  - **a ball from the first box with probability $\frac{1}{36}$;**
  - **a ball from the second box with probability $\frac{2}{36}$;**
  - **a ball from the third box with probability $\frac{3}{36}$;**
  - **a ball from the fourth box with probability $\frac{4}{36}$;**
  - **a ball from the fifth b0x with probability $\frac{5}{36}$;**
  - **a ball from the sixth b0x with probability $\frac{6}{36}$;**
  - **a ball from the seventh box with probability $\frac{5}{36}$;**
  - **a ball from the eighth box with probability $\frac{4}{36}$;**
  - **a ball from the ninth box with probability $\frac{3}{36}$;**
  - **a ball from the tenth box with probability $\frac{2}{36}$;**
  - **a ball from the eleventh box with probability $\frac{1}{36}$;**

The METROPOLIS ALGORITHM

This is a general method to transform any irreducible Markov chain on a state space Ω to a time-reversible Markov chain with a required probability distribution.

Let us assume that we have already designed an irreducible state space for our Markov chain and now we want to construct a Markov chain on this state space with a stationary distribution $\pi_x = b(x)/B$, where $b(x) > 0$ for all $x \in \Omega$ and $B = \sum_{x\in\Omega} b(x)$ is finite.

Theorem For a finite state space Ω, neighbourhood structure $\{N(x) \,|\, x \in \Omega\}$, $N = \max_{x\in\Omega} |N(x)|$ let $M \geq N$ be any such number. For any $x \in \Omega$, let π_x be the desired probability of state x in the stationary distribution. Consider a Markov chain with

$$P_{x,y} = \begin{cases} \frac{1}{M} \cdot \min\{1, \pi_y/\pi_x\} & \text{if } x \neq y \in N(x) \\ 0 & \text{if } x \neq y \notin N(x) \\ 1 - \sum_{y\neq x} P_{x,y} & \text{if x=y} \end{cases}$$

Then, if this chain is irreducible and aperiodic, the stationary distribution is given by the probability π_x.

COMMUTE and COVER TIME

The commute time C_{uv} between u and v is defined as

$$C_{uv} = h_{uv} + h_{vu} = C_{vu},$$

and it is the expected time for a random walk starting at u to return to u after at least one visit to v.

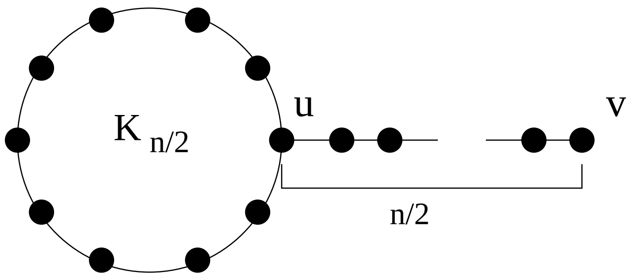
$C_u(G)$ denotes the expected length of a walk that starts at u and ends upon visiting every node in G at least once.

The cover time of G, notation $C(G)$, is defined by

$$C(G) = \max_u C_u(G).$$

EXAMPLE

For the lollipop graph (cukrátko) L_n



it holds

$$h_{uv} = \theta(n^3), \quad h_{vu} = \theta(n^2)$$

(L_n is an example showing that adding more edges can reduce the cover time – contrary to our usual intuition.)

- **A chain of n vertices has the cover time $\theta(n^2)$.**
- **L_n has cover time $\theta(n^3)$.**
- **Complete graph K_n has the cover time $\theta(n \lg n)$.**

COMMUTE TIME - BASIC RESULT

Lemma If $G=(V,E)$, $m=|E|$, $(u,v)\in E$, then $h_{uv}+h_{vu}\leq 2m$.

Proof Let us assign to G a new Markov chain, M_G, states of which are oriented versions of the edges of E ($2m$ of them), and let the transition matrix Q of M_G have only the following non-zero values

$$Q_{(u,v),(v,w)} = p_{vw} = \frac{1}{d(v)}.$$

Matrix Q is **doubly stochastic** (rows and also columns sum-up to 1). Indeed, for each $v,w\in V$:

$$\sum_{x\in V, y\in\Gamma(x)} Q_{(x,y),(v,w)} = \sum_{u\in\Gamma(v)} Q_{(u,v),(v,w)} = \sum_{u\in\Gamma(v)} p_{vw} = d(v)\times\frac{1}{d(v)} = 1.$$

It can be shown, that for any Markov chain with a doubly stochastic matrix the uniform distribution on edges is stationary.

Therefore, (by ergodic theorem) stationary probability of each edge is $\frac{1}{2m}$. Consequently, the expected time between successive traversals of the directed edge (v,u) is $2m = 1/(1/2m)$.

Let us now go back to the original problem $\quad h_{uv}+h_{vu} = ???$

Conditioned on the event that the initial entry into u was through edge (v,u), we can conclude, from the above analysis of M_G, that the expected time to go from there to v and then back to u, along (v,u) is at most $2m$.

The memoryless property of Markov chains allows now to remove conditioning and to get the Lemma.

ELECTRICAL NETWORKS - BASICS

Many random variables associated with simple random walks on undirected graphs can be studied conveniently using concepts and language of the electrical network theory.

A resistive electrical network is an undirected graph where to each edge an edge resistance (as a positive integer) is associated.

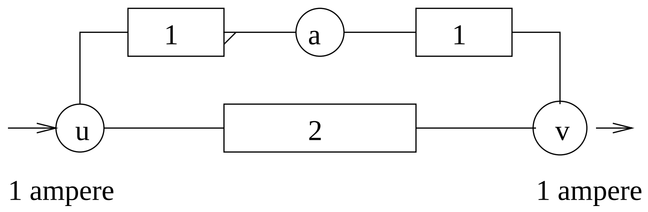


Figure 1: Potentials of nodes are $\phi(u) = 2; \phi(a) = \frac{3}{2}; \phi(v) = 1$; voltage difference between u and v is 1 and between a and v is $\frac{1}{2}$.

Kirchoff law: The sum of all currents entering a node u of a network equals the sum of all currents leaving u.
Ohm law: $I = \frac{V}{R}$, where I is current; V is voltage; R is resistance.

For each edge e, let R_e be the resistance of e. For each node u, let i_u be the current that enters (and exits) the node u.

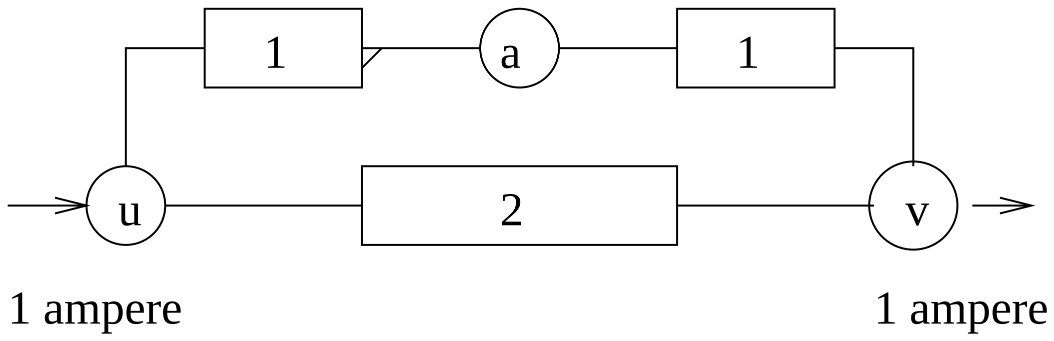


Figure 2: Potentials of nodes are $\phi(u) = 2; \phi(a) = \frac{3}{2}; \phi(v) = 1$; voltage difference between u and v is 1 and between a and v is $\frac{1}{2}$; resistance between u and v is 2; effective resistance is only 1.

We would like to compute the potential $\phi(v)$ for each node v and the current i_{uv} for each edge $e = (u, v)$. By Ohm's Law

$$i_{uv} = \frac{\phi(u) - \phi(v)}{R_e}.$$

Effective resistance R_{uv} between two nodes u and v is the voltage difference between u and v when one ampere is injected into u and removed from v. Hence $R_{uv} = \phi(u) - \phi(v)$.

Observe that effective resistance between nodes of an edge can be smaller than the resistance of the edge (see the above figure).

ELEKTRICKÝ POTENCIÁL a NAPÄTIE

Elektrický potenciál je práca potrebná k preneseniu jednotkového náboja z nekonečna na dané miesto silového poľa po ľubovolnej dráhe.

Potenciál je veličina charakterizujúca energetický stav v danom bode elektrického silového poľa.

Elektrické napätie je práca, ktorá sa vykoná v elektrickom poli pri preneseni jednotkového elektrického množstva po určitej dráhe.

V potenciálovom elektrickom poli nezávisi napätie na dráhe a rovná sa rozdielu potenciálov.

COMMUTE TIME and EFFECTIVE RESISTANCE

To each graph G we associate an electrical network $\mathcal{N}(G)$ at which the resistance of each edge is 1.

Theorem Commute time for any two vertices u and v in G is

$$C_{uv} = 2mR_{uv} = h_{uv} + h_{vu}.$$

Proof: Notation:

- $d(x)$ is the degree of any node x.
- Φ_{zv} is the potential of any node z, relative to v, when $d(x)$ units of current enter each node $x \in V$, and $2m$ units of the current exit v.

For every edge $(u,w) \in E$ by Ohm's law:

$$\Phi_{uv} - \Phi_{wv} = i_{uw} R_{uw} = i_{uw}.$$

Kirchhoff's law implies that for every $u \in V - \{v\}$

$$d(u) = \sum_{w \in \Gamma(u)} i_{uw} = \sum_{w \in \Gamma(u)} \Phi_{uv} - \Phi_{wv}.$$

On the other hand, definition of $h_{uv}, (u,v) \in E(G)$, implies that for each $u \in V - \{v\}$:

$$h_{uv} = \sum_{w \in \Gamma(u)} \frac{1}{d(u)} (1 + h_{wv}).$$

It can be shown, that both above systems of equations are in fact the same system of linear equations and therefore they have the same solution

$$h_{uv} = \Phi_{uv},$$

for each $u, v \in V$.

DERIVATION

Kirchhoff's law implies that for every $u \in V - \{v\}$

$$d(u) = \sum_{w \in \Gamma(u)} i_{uw} = \sum_{w \in \Gamma(u)} \Phi_{uv} - \Phi_{wv}.$$

Hence,

$$d(u) = \Phi_{uv} d(u) - \sum_{w \in \Gamma(u)} \Phi_{wv}$$

and therefore

$$1 = \Phi_{uv} - \frac{1}{d(u)} \sum_{w \in \Gamma(u)} \Phi_{wv}$$

and

$$\Phi_{uv} = 1 + \frac{1}{d(u)} \sum_{w \in \Gamma(u)} \Phi_{wv}$$

and therefore

$$\Phi_{uv} = \frac{1}{d(u)} \sum_{w \in \Gamma(u)} 1 + \frac{1}{d(u)} \sum_{w \in \Gamma(u)} \Phi_{wv} = \sum_{w \in \Gamma(u)} \frac{1}{d(u)} (1 + \Phi_{wv}).$$

and from that one can easily see that we got the same equation as.

$$h_{uv} = \sum_{w \in \Gamma(u)} \frac{1}{d(u)} (1 + h_{wv}).$$

PROOF – CONTINUATION

Consider now the following network: a current of magnitude $2m$ enters u and a current of magnitude $d(x)$ exits every node $x \in V$.

If the potential of u in this network is assumed to be equal to 0, then the potential of v is equal to $-\phi_{uv} = -h_{uv}$.

Let us now perform a superposition of such a network with the network considered on previous slide (at which $d(x)$ units of current enter each node $x \in V$, and $2m$ units of the current exit v).

In the resulting network, all external currents cancel, except for those in vertices u (where the current of magnitude $2m$ enters) and v (where the current of magnitude $2m$ exits).

The difference of potentials between u and v will be

$$h_{uv} - (-h_{vu}) = h_{uv} + h_{vu} = \Phi_{uv} + \Phi_{vu} = C_{uv}.$$

Therefore, C_{uv} is the voltage between u and v in the last network.

Hence, by Ohm law

$$C_{uv} = 2mR_{uv}$$

Claim: **For every pair of vertices u and v, the effective resistance R_{uv} is not more than the distance between u and v in G.**

Corollary: **Let $G = (V, E)$ be a graph, $n = |V|$, $m = |E|$ and $u, v \in V$.**

- **If $(u, v) \in \mathcal{N}(G)$, then $C_{uv} \leq 2m$;**
- **If $u, v \in V$, then $C_{uv} \leq 2m(n-1)$.**
- **If $u, v \in V$, then $C_{uv} < n^3$.**

COVER TIME

Theorem For a graph G with n nodes and m edges $C(G) \leq 2m(n-1)$.

Proof Let T be any spanning tree of $G = (V, E)$. Then there is a traversal of T visiting nodes

$$v_0, v_1, \ldots, v_{2n-2} = v_0$$

that traverses each edge of T exactly once in each direction.

Consider a random walk that starts at v_0, visits all nodes in the order prescribed by the traversal, and terminates after returning to v_0.

An upper bound on the expected length of such a walk is an upper bound on $C_{v_0}(G)$.

$$C_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j v_{j+1}} = \sum_{(u,v) \in T} C_{uv}.$$

Since $(v_j, v_{j+1}) \in E$, $C_{v_j v_{j+1}} \leq 2m$ by previous corollary.

Therefore

$$C_{v_0}(G) \leq 2m(n-1).$$

The above bound is independent of the choice of v_0. Hence

$$C(G) \leq 2m(n-1).$$

SPECIAL CASES of GRAPHS

Let us derive cover time for several special graphs.

- **A complete graph** K_n. The problem to determine cover time is essentially the Coupon Collector problem. Therefore
$$C(K_n) = n \lg n + cn.$$
- **A star graph** S_n The problem of calculating the cover time is again essentially the Coupon Collector problem. Therefore
$$C(S_n) = 2n \lg n + cn.$$
- **A line** L_n**:** Let u_1, u_2 be the end points of L_n. Since $R_{u_1,u_2} = n-1$, we have
$$C_{u_1,u_2} = 2|E(L_n)|R_{u_1,u_2} = 2(n-1)(n-1) = 2(n-1)^2.$$
By symmetry,
$$h_{u_1,u_2} = (n-1)^2.$$
Then
$$C(L_n) = (n-1)^2.$$
- **Lollipop graph** L_n**:** $C(L_n) = \theta(n^3)$.

EFFECTIVE RESISTANCE of GRAPHS

The effective resistance $R(G)$ of a graph G is defined by

$$R(G) = \max_{\{u,v\}\subset V(G)} R_{uv}.$$

Theorem $mR(G) \leq C(G) \leq 2e^3 mR(G)\ln n + n$.
Proof. Lower bound: Let $R(G) = R_{uv}$ for some vertices $u, v \in V$. Then

$$C(G) \geq \max\left(h_{uv}, h_{vu}\right) \geq \frac{C_{uv}}{2} = \frac{2mR_{uv}}{2} = mR(G).$$

Upper bound. Create a random walk of length $2e^3 mR(G)\ln n$ and divide it into $\ln n$ phases of the same length.

For any vertices u and v, the hitting time h_{uv} is at most $2mR(G)$. (This is the average time to get through any of $\ln n$ phases.)

By Markov inequality ($Pr[Y \geq t] \leq \frac{\mathbf{E}[Y]}{t}$), the probability that v is not visited during a single phase is at most $\frac{2mR(G)}{2e^3 mR(G)} (= \frac{h_{uv}}{2e^3 mR(G)}) = \frac{1}{e^3}$.

Therefore, the probability that v is not visited during any of the $\ln n$ phases is at most $(\frac{1}{e^3})^{\ln n} = \frac{1}{n^3}$.

Summing over n choices of v, we get that the probability that there is a node not visited within $2e^3 mR(G)\ln n$ steps is at most $\frac{1}{n^2}$.

When this happens (that is if there is a node not visited during $2e^3 mR(G)\ln n$ steps), we "continue to walk until all nodes are visited" (and n^3 steps suffices for that).

The expected total time is therefore

$$2e^3 mR(G)\ln n + (\frac{1}{n^2})n^3 = 2e^3 mR(G)\ln n + n.$$

APPLICATION of RAYLEIGHT's MONOTONICITY LAW

Rayleight's monotonicity law states that the effective resistance of a graph is non-increased (non-decreased), whenever the resistance of any edge of the graph is decreased (increased).

As a corollary, the effective resistance of the graph can not increase by adding more edges.

Lemma The effective resistance of a graph is not more than its diameter diam(G).

Proof The whole graph can be generated by adding edges to the subgraph that corresponds to the diameter.

Fact: If G is a k-regular graph with n edges, then diam$(G) \leq \frac{3n}{k}$.

Theorem If G is a k-regular graph with n edges, then $C(G) = \mathcal{O}(n^2 \ln n)$.

Proof. Since

$$n \geq \frac{k \cdot \text{diam}(G)}{3},$$

and, by the last theorem, $C(G) = \mathcal{O}(mR(G) \ln n)$, we have $R(G) \leq$ diam$(G) \leq \frac{3n}{k}$ and

$$C(G) \leq \mathcal{O}(\frac{nk}{2} \cdot \frac{3n}{k} \cdot \ln n).$$

USTCON PROBLEM

It is the problem to decide, given an undirected graph G and two vertices s and t, whether there is a path from s to t.

Let **RLP** be the family of languages L for which there exists a probabilistic off-line log-space TM M such that for any input x

$$Pr[\mathcal{M} \text{ accepts } x] \begin{cases} \geq \frac{1}{2} & \text{if } x \in L \\ = 0 & \text{if } x \notin L \end{cases}$$

Theorem USTCON $\in$ **RLP**.

Proof Let a log-space bounded probabilistic TM $\mathcal{M}$ simulate a random walk of length $2n^3$ through the given graph starting from s.

If $\mathcal{M}$ encounters t during such a walk, it outputs YES, otherwise it outputs NO. The probability of the output YES instead of NO is 0.

What is the probability that $\mathcal{M}$ outputs NO instead of YES?

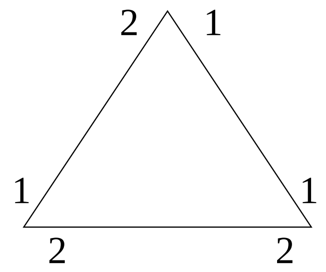
We know that $h_{st} \leq n^3$. By Markov inequality, if t is reachable from s, then the probability that t is not visited during $2n^3$ steps is at most $\frac{1}{2}$.

$\mathcal{M}$ needs a space to count till $2n^3$ and to keep track of its position in the graph during the walk. Therefore it needs space

$$\mathcal{O}(\lg n).$$

Nonuniform, deterministic, log-space algorithms for USTCON

We will consider regular d-degree graphs with n nodes such that all edges of each node are labeled by labels from $\{1, 2, \ldots, d\}$. Any $\sigma \in \{1, 2, \ldots, d\}^*$ or $\sigma \in \{1, 2, \ldots, d\}^\infty$ and any starting node v



specify a walk through the given graph.

A sequence σ is said to **traverse** a graph G if the walk it prescribes visits every node of G regardless of the starting node.

A sequence σ is said to be a **universal traverse sequence** for a class of labeled graphs if it traverses every labeled graph in the class (for any starting node).

A universal traversal sequence whose length is polynomial in n can be used by a deterministic log-space off-line TM to decide instances of USTCON.

(However, in order to be a uniform log-space algorithm, the universal traversal sequence should be constructable by a log-space TM, rather than be encoded in the machine's finite state control.)

UNIVERSAL TRAVERSAL SEQUENCE

$\mathcal{G}$ – a family of connected regular graphs on n-nodes and m edges.

$U(\mathcal{G})$ — length of the shortest universal traversal sequence for $\mathcal{G}$.
$R(\mathcal{G})$ — maximum resistance between any two nodes of any graph in $\mathcal{G}$,

Theorem $U(\mathcal{G}) \leq 5mR(\mathcal{G})\lg(n|\mathcal{G}|)$.

Proof Given $G \in \mathcal{G}$, $v \in G$, let us consider a random walk of the length

$$5mR(\mathcal{G})\lg(n|\mathcal{G}|)$$

divided into $\lg(n|\mathcal{G}|)$ sections of length $5mR(\mathcal{G})$.

The probability that the walk fails to visit v in any section is at most $\frac{2}{5}$. (Due to the Markov inequality and the fact that $C_{uv} = 2mR_{uv}$.)

Therefore, the probability that v is **not** visited during any of the $\lg(n|\mathcal{G}|)$ sections is thus at most

$$(\frac{2}{5})^{\lg(n|\mathcal{G}|)} = (\frac{2}{5})^{(\lg_{2/5}(n|\mathcal{G}|)/\lg(2/5)} = (n|\mathcal{G}|)^{1/(\lg(2/5)} = (n|\mathcal{G}|)^{-c} \text{ for a } c > 1.$$

Summing up over n choices of v and $|\mathcal{G}|$ choices of the labeled graph G, the probability that the random walk (sequence) fails to be universal is less than 1.

As a consequence, there is a sequence of such a length that is universal for the class $\mathcal{G}$. (We have just used the probabilistic method.)

CONVERGENCE of RANDOM WALKS

In the rest of chapter we deal with the following problem: how many steps are needed for a random walk (that starts at some node), to converge to the stationary distribution - that is to be in each node with probability specified by the stationary distribution.

We present two techniques to deal with the above problem.

Stopping rule method calculates the rate of convergence directly by defining a proper stopping rule.

Coupling method reduces the problem of determining the rate of convergence to that of calculating the number of steps needed to meet another, imaginary, random walk, that starts at the stationary distribution.

PRELIMINARIES

Variation distance, $||P-Q||$, between two probability distributions, P and Q, on a set I of states is defined by

$$||P-Q|| = \max_{I' \subseteq I} |P(I') - Q(I')|.$$

Lemma Let P and Q be probability distributions over a set I of states. then

$$||P-Q|| = \frac{1}{2}\sum_{i \in I} |P(i) - Q(i)|.$$

STOPPING RULE (TIME)- REPETITION

Given is a sequence of random variables $Z_1, Z_2, \ldots$, and another random variable T, whose range are natural numbers. T is a **stopping rule** for random variables $Z_1, Z_2, \ldots$, if, for every i, the event $T = i$ is independent of variables Z_j, for all $j > i$.

The idea is that the variables Z_i are observed in order one at each time moment: at first Z_1, then Z_2 and so on. The value of the variable T then shows the number of variables observed when we stop.

It is sometimes required that $Pr(T < \infty) = 1$, or that T is almost surely finite.

The intuition behind such a definition of the stopping rule is that at any particular time it is enough to look at the sequence so far in order to be able to tell if it is time to stop.

STOPPING RULE - OBSERVATIONS

Stopping rule can be seen as a mechanism for deciding whether to continue or to stop a process on the basis of the present position and past events, and which will always lead to a decision to stop at some time.

Examples Consider a gambler playing roulette, starting with \$ 100.

- Playing one, and only one, game corresponds to the stopping time $T = 1$, and this is a stopping rule.
- Playing until she either runs out of money or has played 500 games is a stopping rule.
- Playing until she doubles her money is not a stopping rule (here it is assued that betting systems has some limitations on doubling or tripling the money).
- Playing until she either doubles her money or runs out of money is a stopping rule (because probability that she stops in finite time is 1)..
- Playing until she is the maximum amount ahead she will ever be is not a stopping rule (as it requires information about future, present and past).

STRONG UNIFORM STOPPING RULE

Given are random variables $Z_1, Z_2, \ldots$, and a random variable T, whose range are natural numbers. T is a stopping rule for variables $Z_1, Z_2, \ldots$, if, for every i, the event $T = i$ is independent of variables Z_j, for all $j > i$.

For a finite ergodic Markov chain, a strong uniform stopping time T is a stopping rule which satisfies the condition

$$\mathbf{Pr}[Z_k = i \,|\, T = k] = \pi_i,$$

where Z_k is the state at the kth step in the Markov chain, and π_i is the probability, under the stationary distribution, of being at the state i.

Next theorem relates strong uniform stopping rule (time) and the rate of convergence of the random walk.

Theorem **Let π be the stationary distribution of a random walk, and $Q^{(t)}$ be the probability distribution after t steps. In addition, let T be a strong uniform stopping time. Then**

$$||Q^{(t))} - \pi|| \leq \mathbf{Pr}[T > t].$$

Proof Let X_t be random variable producing states from I visited by random walk at step t. It holds

$$\begin{aligned}
\forall I' \subseteq I, Q^{(t)}(I') &= \Pr[X_t \in I'] \\
&= (\sum_{j \leq t} \Pr[(X_t \in I') \cap (T = j)]) + \Pr[(X_t \in I') \cap (T > t)] \\
&= \sum_{j \leq t} \Pr[X_t \in I' | T = j] \Pr[T = j] \\
&+ \Pr[X_t \in I' | T > t] \Pr[T > t] \\
&= \sum_{j \leq t} \pi(I') \Pr[T = j] + \Pr[X_t \in I' | T > t] \Pr[T > t] \\
&= \pi(I')(1 - \Pr[T > t]) + \Pr[X_t \in I' | T > t] \Pr[T > t] \\
&\leq \pi(I') + \Pr[T > t] \Pr[X_t \in I' | T > t] \leq \pi(I') + \Pr[T > t]
\end{aligned}$$

Third equality from the end follows from the fact that T is a strong uniform stopping time and that if a random walk is in a stationary distribution, then it stays in it forever.

The last but one equality follows from the fact that

$$\sum_{j \leq t} \mathsf{Pr}[T = j] = \mathsf{Pr}[T \leq t] = 1 - \mathsf{Pr}[T > t].$$

Last equality implies, since both $\mathsf{Pr}[X_t \in I' | T > t]$ and $\pi(I')$ are at most 1,

$$\forall I' \subseteq I, |Q^{(t)}(I') - \pi(I')| \leq \mathsf{Pr}[T > t].$$

EXAMPLE – HYPERCUBE

We show how fast converges to the stationary distribution a random walk on a hypercube.

Consider the following random walk

1. Choose uniformly a neighbor (of currently visited node).
2. With probability $\frac{1}{2}$ move to the chosen node; otherwise do not move at all.

(Last trick is needed in order to have ergodic (aperiodic) Markov chain.)

Let us define the stopping rule T to be the number of coordinates **chosen** so far (even if not all choices of coordinates yielded a move).

Note that T is a strong uniform stopping rule (informally said, this happens because we have an equal probability of being at any node after all coordinates are chosen).

In order to be able to use the last theorem we need to determine $\Pr[T > t]$. However, this is exactly the coupon selector problem, since we can see coordinates as being coupons that need to be collected.

For coupon selector problem, the expected number of trials needed to see all n coupons is $\mathcal{O}(n \lg n)$. Therefore, the expected number of steps until we choose all coordinates is $\mathcal{O}(n \lg n)$.

Using Markov's inequality we get

$$\Pr[T > t] \leq \frac{\mathbf{E}[T]}{t} = \mathcal{O}(\frac{n \lg n}{t}).$$

EXAMPLE – CARD SHUFFLING I.

Suppose we want to shuffle a pack of n cards, numbered $1, 2, \ldots, n$, according to the following policy: *Move the top card to a random location in the pack.*

How long it will take to have all cards in a random distribution (to reach the stationary distribution)?

This problem can be viewed as a random walk in a graph with $n!$ vertices corresponding to all possible permutations. Edges are determined by the shuffling policy.

Denote by BOTTOM the card that was originally at the bottom of the pack.

Let T be the number (name) of the card moved from the top at the last step. T is stopping rule and $T =$ BOTTOM is the stopping time.

We claim that T is strong uniform time. This is due to the fact that once we remove BOTTOM from the top of the pack, we are already in the stationary distribution. (It can be shown by induction on the number of cards bellow BOTTOM that the cards below BOTTOM are always in a random order.)

We show now that the above stopping rule behaves as a coupon collector.

Indeed, define T_i to be the number of steps until there are i cards bellow BOTTOM (including BOTTOM). Since $T = T_n$ we have

$$T = T_1 + (T_2 - T_1) + (T_3 - T_2) + \ldots + (T_n - T_{n-1}).$$

Moreover, $T_{i+1} - T_i$ has a geometric distribution with parameter $\frac{n-i}{n}$.

This is similar to the situation in coupon selection. Indeed, let V_i denote the number of steps until we see i coupons and let $V = V_n$. Similarity between V and T follows from the fact that

$$V = V_1 + (V_2 - V_1) + (V_3 - V_2) + \ldots + (V_n - V_{n-1})$$

and that $V_{i+1} - V_i$ has also geometric distribution with parameter $\frac{n-i}{n}$. Hence

$$\Pr[T > t] \leq \frac{\mathbf{E}[T]}{t} = \mathcal{O}\left(\frac{n \lg n}{t}\right).$$

COUPLING METHOD – BASIC IDEA

The goal is again to investigate how fast a random walk X, starting at a fixed point, approaches the stationary distribution.

To do that we consider another random walk Y that starts at the stationary distribution.

Each of these two walks is independent from its own point of view.

Y always remains at the stationary distribution.

These two walks are correlated in the sense that once they meet, their future moves are the same.

Therefore, in order to determine how fast X approaches the stationary distribution, it is sufficient to determine when such two random walks meet.

EXAMPLE I – HYPERCUBE II.

Let coordinator be a super-player that coordinates moves of two walks X and Y, at each step, as follows.

1. **The coordinator chooses randomly an $i \in \{1, \ldots, n\}$. Let X_i (Y_i) be the ith bit of the node currently visited by $X(Y)$.**
2. **If $X_i = Y_i$, then with probability $\frac{1}{2}$ both X and Y move, and with probability $\frac{1}{2}$ both of them stay still. The move, if any, is to the neighbor that differs in the ith bit from the node currently visited.**
3. **If $X_i \neq Y_i$, then with probability $\frac{1}{2}$ player X moves and Y stay still, and with the same probability it is vice verse. (Move is again to the node that differs in the ith coordinate.)**

From the point of view of both, X and Y, they perform a random walk.

It is now easy to see that the following claim holds

If, for some i, $X_i = Y_i$, then it always stays as such. If a coordinate i is chosen, (and $X_i \neq Y_i$), then $X_i = Y_i$ at the end of the step.

Hence, the random variable that counts the number of steps needed for X and Y to meet behaves as a coupon collector.

This way we get the same result as using the previous method.

EXAMPLE II – CARD SHUFFLING II.

We will consider again an n *cards shuffling problem*, but this time with a different shuffling policy.

> **Uniformly and randomly choose a card and put it, alternatively, either to the top or to the bottom of the pack.**

To demonstrate *the coupling argument* we will consider two initial packs of cards. One that is fully ordered (first is the card number 1, last the card number n), second that is randomly ordered.

The *coordinator* chooses, randomly, a card i and in both packs the card is moved alternatively to top or to bottom.

Claim The stationary distribution is reached after all cards are picked.

By induction, one can show that after i steps each pack can be partition into three parts: top, middle and bottom.

The cards in the top and bottom parts have all already been selected and appear in the same order in both packs. The cards in the middle part have not been selected yet and appear in some arbitrary order in both packs.

Therefore, after all cards have been picked at least once, the two packs are in the same order.

Observe now that the problem of picking all cards is again similar to the coupon selector problem. The expected number of steps needed to shuffle the pack is therefore $\mathcal{O}(n \lg n)$.

FIXING A SPANNING TREE

For a graph $G = (V, E)$ we want to generate randomly a spanning tree of G. (To make the exposition simple, we assume that edges of the spanning tree will always be oriented towards a specific node r, called the *root* of the tree.)

Consider a Markov chain the states of which are all possible pairs (T, r), where T is a spanning tree of G, and r is the root of T. (The number of states in this Markov chain is $s|V|$, where s is the number of spanning trees of G.) A random walk on this Markov chain will now be defined as follows:

1. **Randomly choose an edge $e \in E$ that connects the root r to a node y;**
2. **If the edge e is in the tree (and therefore directed from y to r), then change its direction - making by that y to be the root of the (new) tree;**

3. If e is not in the tree, add e to the tree and direct it from r to y and delete from the tree the (unique) outgoing edge leaving y.

Observe:

- The indegree of each state of the Markov chain is equal to the degree of the root of the corresponding tree (in the graph) - because a state (T, r) can be reached from all neighbours of r in G;
- The outdegree of each state is equal to the degree of the root of the corresponding tree (in the graph) - argument is the same as in the previous case;
- As a consequence, the probability of choosing a state (T, r) at the stationary distribution is proportional to the degree of r in G.

We use now coupling argument to find out how fast the random walk approaches the the stationary distribution.

Consider two random walks: first walk starts at the stationary distribution; second starts at the arbitrary state.

Stage 1 Two walks will progress independently until they have the same node as the root (observe that trees might still be different).

Stage 2 Once both walks agree on a root, from now on, they will make the same moves.

Observe that if we imagine the root of the tree as a particle, then it (implicitly) performs a random walk in the graph G - this observation is crucial for the following analysis.

Let us now calculate the expected number of steps until two random walks meet - that is that they generate the same spanning tree.

Calculating the expected number of steps in Stage 1 is equivalent to the following problem: Two particles are moving randomly in a graph. What is the expected number of steps they meet?

Lemma The expected number of steps until two particles meet is at most twice the cover time of the graph.

Concerning the expected number of steps in Stage 2, observe first that at the beginning of this stage both spanning trees (in the corresponding states of the two random walks), have the same root.

Observe that when a new root is chosen, there is an edge connecting

the old root to the new root, and it exists in both trees, implying that both spanning trees remain in the same root.

Suppose the root node is switched from r to r'. Notice that, from now on, the (unique) outgoing edge of node r, will remain the same in both trees. This happens, since the outgoing edge of a node r changes when either r becomes the root, or when it ceases to be the root.

It follows, that a sufficient condition for the two trees to be identical, is that each node in the graph is a root of the tree at some point in time.

This implies that an upper bound on the time for the two trees to converge at Stage II is the cover time of G.

Conclusion: the expected number of steps, in both stages, until the spanning trees become identical, depends linearly on the cover time of the graph.

APPENDIX

SECRETARY PROBLEM

The problem can be formulated as follows:

- There is a single secretariat position to fill.
- There are n applicants for the position, and the value of n is known.
- Each applicant has a unique "quality value" - the interview making committee has no knowledge of those that have not been interviewed yet and no knowledge how large is the best quality value.
- The applicants are interviewed in a random order.
- After each interview, the applicant is immediately accepted or rejected.
- The decision to accept or reject an applicant can be based only on the relative "quality value" of the applicants interviewed so far.
- Rejected applicants cannot be recalled.
- The goal is to select an applicant with the best 'quality value". The payoff is 1 for the best applicant and 0 otherwise.

Terminology: A candidate is an applicant who, when interviewed, is better than all the applicants interviewed previously.

SOLUTION

Since the goal in the secretary problem is to select the single best applicant, only candidates will be considered for acceptance.

Optimal policy for this problem (the stopping rule) has a surprising feature: For large n the optimal policy is to interview and reject the first $\frac{n}{e}$ applicants and then to accept the next one who is better than candidates interviewed till then.

As n gets larger, the probability of selecting the best applicant goes to $\frac{1}{e}$, which is around 37%.

ANDREY ANDREEVITCH MARKOV

- **Russian mathematician (1856-1922)**
- **Introduced Markov chains in 1906**
- **The original motivation was to extend the law of large numbers to dependent events.**
- **In 1913 he applied his findings to the first 20 000 letters of Pushkin's Eugene Onegin**

APPLICATIONS of MARKOV CHAINS

- Physics - especially thermodynamics and statistical mechanics
- Chemistry
- The PageRank of a webpage as used by Google is defined by a Markov chain. It is the probability to be at page i in the stationary distribution on the following Markov chain on all known webpages.

  If N is the number of known webpages, and a page i has links to k_i webpages, then the probability to go to any of these pages is
  $$\frac{\alpha}{k_i} + \frac{1-\alpha}{N}$$
  the and probability to go to any other page is
  $$\frac{1-\alpha}{N}$$
  The parameter α is about 0.85.

- **Economics and finance**
- **Social sciences**
- **Mathematical biology**
- **Algorithmic music composition**