Probabilistic Aspects of Computer Science: TD2
Markov chains in the long run

Adrien Koutsos

September 19, 2017

Exercise 1. We study again the same exercise than last week, but with simpler tools.

1. Let X_n be the number of heads obtained after n independent tosses of a (possibly unfair) coin. Show that, for any $k \geq 2$,
$$\lim_{n\to\infty} \mathbf{Pr}(X_n \text{ is divisible by } k) = \frac{1}{k}$$
2. Solve the problem when X_n represents the sum of n independent rolls of a dice.

Exercise 2. Exhibit a Markov chain which has null recurrent states (different from the one studied in course).

Exercise 3. Show that if a Markov chain has two steady-state distributions, then it has an infinite number of steady-state distributions.

Exercise 4 (Move-to-front heuristic)**.** Suppose that we are given $n \geq 2$ records $R_1, R_2, \ldots, R_n$. The records are kept in some order. The cost of accessing the jth record in the order is j. Thus, if we had four records ordered as R_2, R_4, R_3, R_1, then the cost of accessing R_4 would be 2 and the cost of accessing R_1 would be 4.

Suppose further that, at each step, record R_j is accessed with probability p_j, with each step being independent of other steps.

1. If we knew the values of the p_j in advance, what is the best choice to order the records?

We suppose now that we do not know the p_j in advance and we use a *move-to-front* heuristic: at each step, put the record that was accessed at the front of the list. We assume that moving the record can be done with no cost and that all other records remain in the same order. For example, if the order was R_2, R_4, R_3, R_1 before R_3 was accessed, then the order at the next step would be R_3, R_2, R_4, R_1.

3. Describe this problem with a Markov chain whose set of states is the set of permutation over n elements, and show that it admits a steady-state distribution.
4. We define
$$\varphi_n(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{x_i}{\sum_{j=i}^{n} x_j}$$
Show that for all $\pi_\sigma = \varphi_n(p_{\sigma(1)}, \ldots, p_{\sigma(n)})$ is the steady-state distribution of this chain.
5. Let X_k be the cost for accessing the kth requested record, $V_{k,i}$ be the event that the k-th query refers to record R_i. Using the law of total expectation we get that:
$$\mathbf{E}(X_k) = \sum_{i=1}^{n} \mathbf{E}(X_k \mid V_{k,i}) \mathbf{Pr}(V_{k,i})$$
   (a) Let $V'_{k,i}$ the event that the kth query refers to R_i and some query previous to the kth also referred to R_i. Show that:
$$\lim_{k\to+\infty} \mathbf{E}(X_k \mid V_{k,i}) = \lim_{k\to+\infty} \mathbf{E}(X_k \mid V'_{k,i})$$

(b) Let $M_{k,i}$ be the number of queries between the k-th query and the last time record R_i was queried. Using that, give an expression, without probabilities, of $\mathbf{E}(X_k \mid V'_{k,i})$.

(c) Conclude by showing that:

$$\lim_{k\to\infty} \mathbf{E}(X_k) = \frac{1}{2} + \sum_{i,j} \frac{p_i p_j}{p_i + p_j}\,.$$

Exercise 5 (Cover time)**.** Let $G = (V, E)$ be a finite, undirected, and connected graph. A random walk on G is a Markov chain defined by the sequence of moves of a particle between vertices of G. In this process, the place of the particle at a given time step is the state of the system. If the particle is at vertex i and if i has $d(i)$ outgoing edges, then the probability that the particle follows the edge $\{i, j\}$ and moves to a neighbor j is $1/d(i)$.

1. Show that a random walk on an undirected graph G is aperiodic if and only if G is not bipartite.

2. In the rest of the exercise, we assume that G is not bipartite. Show that a random walk on G converges to a steady-state distribution π, where $\pi_v = \frac{d(v)}{2|E|}$.

3. We consider a new Markov chain defined on the edges of G. The current state is defined to be the pair composed of the edge most recently traversed in the random walk, together with the direction of this traversal: the state space is hence the set of directed edges. There are $2|E|$ states in this new Markov chain, and its transition matrix $\mathbf{Q}$ is given by:

$$\mathbf{Q}_{(u,v),(v,w)} = \begin{cases} \frac{1}{d(v)} & \text{if } \{u, v\} \in E \\ 0 & \text{otherwise} \end{cases}$$

   Compute its steady-state distribution.

4. We denote $\mu_{v,u}$ the expected number of steps to reach u from v. Show that if $\{u, v\} \in E$, then $\mu_{u,v} + \mu_{v,u} \leq 2|E|$ *(Hint: use the result of the previous question)*.

5. The *cover time* of G is defined as the maximum over all vertices $v \in V$ of the expected time to visit all of the nodes in the graph by a random walk starting from v. Show that the cover time of G is bounded above by $2|E|(|V| - 1)$.

6. As an application, suppose we are given an undirected graph $G = (V, E)$ and two vertices s and t in G, and we want to determine whether there is a path connecting s and t. For simplicity, assume that the graph G has no bipartite connected components. By standard deterministic search algorithms, we can easily solve the problem in linear time, using $\Omega(n)$ space. Show that the following algorithm returns the correct answer with probability $1/2$, and it only errs by returning that there is no path from s to t when there is such a path. What is the time and space complexities of this algorithm? (*Hint: you may use the Markov's inequality, which says for a random variable X and $a > 0$ that* $\mathbf{Pr}(|X| \geq a) \leq \frac{\mathbf{E}(|X|)}{a}$.)

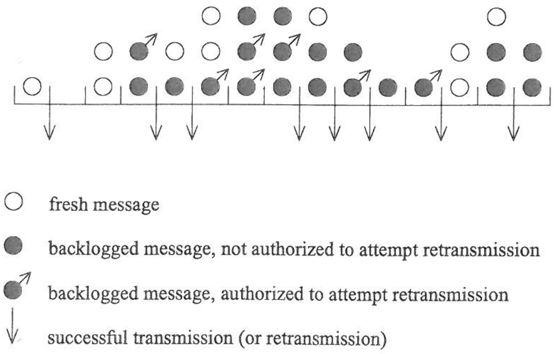
s-t Connectivity algorithm
1. Start a random walk from s.
2. If the walk reached t within $2|V|^3$ steps, return that there is a path. Otherwise, return that there is no path.

Exercise 6 (ALOHA)**.** A typical situation in a multiple-access satellite communications system is the following. Users, each one identified with a message, contend for access to a single-channel satellite communications link for the purpose of transmitting messages. Two or more messages in the air at the same time jam each other, and are not successfully transmitted. The users are somehow able to detect a collision of this sort and will try to retransmit later the message involved in a collision. The difficulty in such communications systems resides mainly in the absence of cooperation among users, who are all unaware of the intention to transmit of competing users. The slotted ALOHA protocol imposes on the users the following rules:

(i) Transmissions and retransmissions of messages can start only at equally spaced moments; the interval between two consecutive (re-)transmission times is called a *slot*; the duration of a slot is always larger than that of any message.

(ii) All *backlogged* messages, i.e., those messages having already tried unsuccessfully – maybe more than once – to get through the link, require retransmission independently of one another with probability $\nu \in (0,1)$ at each slot. This is the so-called *Bernoulli retransmission policy.*

(iii) The *fresh messages* – those presenting themselves for the first time – immediately attempt to get through.



Let X_n be the number of backlogged messages at the beginning of slot n.

1. Supposing there are $X_n = k$ backlogged messages, express the probability $b_i(k)$ that i among them attempt to retransmit in slot n as a function of i, k and ν.

2. Let A_n be the number of fresh requests for transmission in slot n. The sequence $\{A_n\}_{n\geq 0}$ is assumed i.i.d. with the distribution $\mathbf{Pr}(A_n = j) = a_j$. Give a condition over the sequence $(a_j)_{j\geq 0}$ such that the sequence $\{X_n\}_{n\geq 0}$ is described by an irreducible Markov chain.

3. Show that this chain is not positive recurrent: we say that the system using the Bernoulli retransmission policy is *not stable* (*Hint: one may show that a steady state distribution* π *would satisfy* $\lim_{N\to\infty} \frac{\pi_{N+1}}{\pi_N} = \infty$).

   In the following, we admit and use the *Pakes' Lemma*:

   > Let $\{X_n\}_{n\geq 0}$ be an irreducible Markov chain with states $\mathbb{N}$, such that for all $i, n \geq 0$
   > $$\mathbf{E}(X_{n+1} \mid X_n = i) < \infty$$
   > and
   > $$\limsup_{i\to\infty} \mathbf{E}(X_{n+1} - X_n \mid X_n = i) < 0\,.$$
   > Then the Markov chain is positive recurrent.

4. We now consider a retransmission policy stabilizing ALOHA. Assume the retransmission probability ν now depends on the number k of backlogged messages. Express the expectations appearing in Pakes' Lemma as a function of $\nu(k)$, a_0, a_1 and $\lambda \stackrel{\text{def}}{=} \mathbf{E}(A_n) = \sum_{i=1}^{\infty} i a_i$ (the so-called *traffic intensity*, supposed finite from now on).

5. Using Pakes' Lemma, design a $\nu(i)$ and find a sufficient condition over λ, a_0 and a_1 for stability of this protocol.

6. Supposing that the arrivals $\{A_i\}$ follow a Poisson distribution of parameter λ
   $$a_i = e^{-\lambda}\frac{\lambda^i}{i!}\,,$$
   find a condition over λ for the ALOHA protocol to be stable.