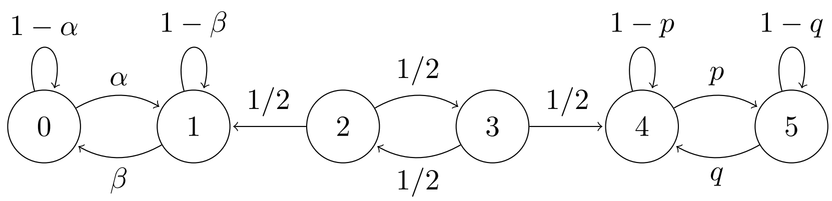
UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 4
Fall 2021

1. Reducible Markov Chain

Consider the following Markov chain, for $\alpha, \beta, p, q \in (0,1)$.



(a) Find all the recurrent and transient classes.

(b) Given that we start in state 2, what is the probability that we will reach state 0 before state 5?

(c) What are all of the possible stationary distributions of this chain? *Hint:* Consider the recurrent classes.

(d) Suppose we start in the initial distribution $\pi_0 := \begin{bmatrix} 0 & 0 & \gamma & 1-\gamma & 0 & 0 \end{bmatrix}$ for some $\gamma \in [0,1]$. Does the distribution of the chain converge, and if so, to what?

Solution:

(a) The classes are $\{0,1\}$ (recurrent), $\{4,5\}$ (recurrent), and $\{2,3\}$ (transient).

(b) Let T_0 and T_5 denote the time it takes to reach states 0 and 5 respectively. (Note that exactly one of T_0 and T_5 will be finite.) We are looking to compute $P_2(T_0 < T_5)$, and we can set up hitting equations:

$$
\begin{aligned}
P_2(T_0 < T_5) &= \frac{1}{2} + \frac{1}{2} P_3(T_0 < T_5), \\
P_3(T_0 < T_5) &= \frac{1}{2} P_2(T_0 < T_5).
\end{aligned}
$$

Thus, $P_2(T_0 < T_5) = 2/3$.

(c) First we observe that no stationary distribution can put positive probability on a transient state, so the stationary distribution is supported on the states $\{0,1,4,5\}$. Next, if we restrict our attention to only the states $\{0,1\}$, then we have an irreducible Markov chain with stationary distribution

$$
\pi_1 := \frac{1}{\alpha+\beta} \begin{bmatrix} \beta & \alpha \end{bmatrix},
$$

and similarly, if we restrict our attention to only the states $\{4, 5\}$, then again we have an irreducible Markov chain with stationary distribution

$$\pi_2 := \frac{1}{p+q} \begin{bmatrix} q & p \end{bmatrix}.$$

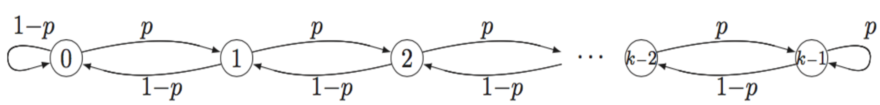
Any stationary distribution for the entire chain must be some convex combination of these two stationary distributions. Explicitly, the stationary distributions are of the form

$$\pi = \begin{bmatrix} \frac{c\beta}{\alpha+\beta} & \frac{c\alpha}{\alpha+\beta} & 0 & 0 & \frac{(1-c)q}{p+q} & \frac{(1-c)p}{p+q} \end{bmatrix} \tag{1}$$

for some $c \in [0, 1]$.

(d) Indeed the distribution will converge, even though we do not have irreducibility. The intuition is as follows. The probability will leak out of the transient states $\{2, 3\}$ until all of the probability mass is supported on the recurrent states. The two recurrent classes can each be considered to be an irreducible aperiodic Markov chain and so the probability mass which enters a recurrent class will settle into equilibrium. To aid us in finding the limiting distribution, we can use the results of Part (b). With probability γ, we start in state 2, and with a further probability 2/3 we end up in the recurrent class $\{0, 1\}$. By symmetry, the probability that we end up in $\{0, 1\}$ starting form state 3 is 1/3. Thus, the total probability mass which settles into the recurrent class $\{0, 1\}$ is $2\gamma/3 + (1-\gamma)/3 = 1/3 + \gamma/3$. Then, the probability mass settling in the recurrent class $\{4, 5\}$ is $2/3 - \gamma/3$. Therefore, the chain converges to the stationary distribution in (**??**) with $c = 1/3 + \gamma/3$.

2. Finite Random Walk



Assume $0 < p < 1$. Find the stationary distribution. *Hint:* Let $q = 1 - p$ and $\rho = \frac{p}{q}$, but be careful when $\rho = 1$.

Solution:

In the stationarity, the net flow of probability through every cut of the graph must be 0. What this means is that the probability flowing from state i to state $i + 1$ must equal the reverse.

$$\pi_i p = q\pi_{i+1}$$

This means that $\pi_{i+1} = \frac{p}{q}\pi_i = \rho\pi_i$, and more generally, that

$$\pi_i = \rho^i \pi_0$$

We can then solve for π_0 since we know that $\sum_{i=0}^{k-1} \pi_i = 1$.

$$\pi_0 \left(\sum_{j=0}^{k-1} \rho^j \right) = 1 \quad \text{so} \qquad \pi_0 = \frac{1-\rho}{1-\rho^k}; \quad \pi_j = \rho^j \frac{1-\rho}{1-\rho^k}.$$

This is derived from the formula for the sum of a geometric series. However, in the special case where $\rho = 1$, the formula is undefined, and the answer is just $\pi_j = 1/k$ for $j = 0, \ldots, k - 1$.

3. Soliton Distribution

This question pertains to the **fountain codes** introduced in Lab 2.

Say that you wish to send n chunks of a message, $X_1, \dots, X_n$, across a channel, but alas the channel is a **packet erasure channel**: each of the packets you send is erased with probability $p_e > 0$. Instead of sending the n chunks directly through the channel, we will instead send n packets through the channel, call them $Y_1, \dots, Y_n$. How do we choose the packets $Y_1, \dots, Y_n$? Let $p(\cdot)$ be a probability distribution on $\{1, \dots, n\}$; this represents the **degree distribution** of the packets.

(i) For $i = 1, \dots, n$, pick D_i (a random variable) according to the distribution $p(\cdot)$. Then, choose D_i random chunks among $X_1, \dots, X_n$, and "assign" Y_i to the D_i chosen chunks.

(ii) For $i = 1, \dots, n$, let Y_i be the XOR of all of the chunks assigned for Y_i (the number of chunks assigned for Y_i is called the **degree** of Y_i).

(iii) Send each Y_i across the channel, along with metadata which describes which chunks were assigned to Y_i.

The receiver on the other side of the channel receives the packets $Y_1, \dots, Y_n$ (for simplicity, assume that no packets are erased by the channel; in this problem, we are just trying to understand what we should do in the ideal situation of *no* channel noise), and decoding proceeds as follows:

(i) If there is a received packet Y with only one assigned chunk X_j, then set $X_j = Y$. Then, "peel off" X_j: for all packets Y_i that X_j is assigned to, replace Y_i with $Y_i \,\mathrm{XOR}\, X_j$. Remove Y and X_j (notice that this may create new degree-one packets, which allows decoding to continue).

(ii) Repeat the above step until all chunks have been decoded, or there are no remaining degree-one packets (in which case we declare failure).

In the lab, you will play around with the algorithm and watch it in action. Here, our goal is to work out what a good degree distribution $p(\cdot)$ is.

Intuitively, a good degree distribution needs to occasionally have prolific packets that have high degree; this is to ensure that all packets are connected to at least one chunk. However, we need "most" of the packets to have low degree to make decoding easier. Ideally, we would like to choose $p(\cdot)$ such that at each step of the algorithm, there is exactly one degree-one packet.

(a) Suppose that when k chunks have been recovered ($k = 0, 1, \dots, N-1$), then the expected number of packets of degree d (for $d > 1$) is $f_k(d)$. Assuming we are in the ideal situation where there is exactly one degree-one packet for any k : What is the probability that a given degree d packet is connected to the chunk we are about to peel off? Based on that, what is the expected number of packets of degree d whose degrees are reduced by one after the $(k+1)$st chunk is peeled off?

(b) We want $f_k(1) = 1$ for all $k = 0, 1, \dots, n-1$. Show that in order for this to hold, then for all $d = 2, \dots, n$ we have $f_k(d) = (n-k)/[d(d-1)]$. From this, deduce what $p(d)$ must be, for $d = 1, \dots, n$. (This is called the **ideal soliton distribution**.)

[*Hint*: You should get two different recursion equations since the only degree 1 node at peeling $k+1$ is going to come from the peeling of degree 2 nodes at peeling k, however, for other higher degree d nodes, there will be some probability that some degree d ones will remain from the previous iteration and some probability that they will come from $d+1$ one that will be peeled off]

(c) Find the expectation of the distribution $p(\cdot)$.

In practice, the ideal soliton distribution does not perform very well because it is not enough to design the distribution to work well in expectation.

Solution:

(a) Of the $f_k(d)$ packets with degree d, each packet has probability $d/(n-k)$ (since there are $n-k$ packets remaining) of being connected with the message packet which is peeled off at iteration $k+1$. Thus, by linearity, the answer is $f_k(d)d/(n-k)$.

(b) We certainly need $f_0(1) = 1$ and $1 = f_1(1) = f_0(2) \cdot 2/n$, so $f_0(2) = n/2$. For $k = 0, 1, \ldots, n-1$, we have $1 = f_{k+1}(1) = f_k(2) \cdot 2/(n-k)$, so $f_k(2) = (n-k)/2$.
Proceed by induction. Suppose that for all $d \le d'$, where $d' = 2, \ldots, n-1$, we know that $f_k(d) = (n-k)/[d(d-1)]$. Then, for $k = 0, 1, \ldots, n-d-1$,

$$\begin{aligned}\frac{n-k-1}{d(d-1)} = f_{k+1}(d) &= f_k(d+1)\frac{d+1}{n-k} + f_k(d)\Big(1 - \frac{d}{n-k}\Big) \\ &= f_k(d+1)\frac{d+1}{n-k} + \frac{n-k}{d(d-1)}\Big(1 - \frac{d}{n-k}\Big)\end{aligned}$$

so $f_k(d+1) = (n-k)/[d(d+1)]$.
Note that $f_0(d)$, the expected number of degree-d received packets at the beginning of the algorithm, is exactly $np(d)$, so:

$$p(d) = \begin{cases} \dfrac{1}{n}, & d = 1 \\ \dfrac{1}{d(d-1)}, & d = 2, \ldots, n \end{cases}$$

(c) The expectation is

$$\sum_{d=1}^{n} dp(d) = \frac{1}{n} + \sum_{d=2}^{n} d \cdot \frac{1}{d(d-1)} = \frac{1}{n} + \sum_{d=2}^{n} \frac{1}{d-1} = \sum_{d=1}^{n} \frac{1}{d} \approx \ln n.$$