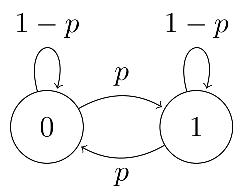
UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 7
Spring 2021

1. Compression of a Markov Chain

Consider an irreducible Markov chain $(X_n)_{n\in\mathbb{N}}$ as shown below.



Suppose $X_0 \sim B(\frac{1}{2})$. Roughly how many bits are needed to represent $(X_0, X_1, \dots, X_n)$?

2. Mutual Information and Channel Coding

The **mutual information** of X and Y is defined as

$$I(X;Y) := H(X) - H(X \mid Y)$$

Here, $H(X \mid Y)$ denotes the **conditional entropy** of X given Y, which is defined as:

$$\begin{aligned} H(X \mid Y) &= \sum_{y\in\mathcal{Y}} p_Y(y) H(X \mid Y = y) \\ &= \sum_{y\in\mathcal{Y}} p_Y(y) \sum_{x\in\mathcal{X}} p_{X|Y}(x \mid y) \log_2 \frac{1}{p_{X|Y}(x \mid y)} \end{aligned}$$

The interpretation of conditional entropy is the average amount of uncertainty remaining in the random variable X after observing Y. The interpretation of mutual information is therefore the amount of information about X gained by observing Y.

The channel coding theorem says that if X is passed into the channel and Y is received, then the capacity of the channel is

$$C = \max_{p_X} I(X;Y) = \max_{p_X} H(X) - H(X \mid Y)$$

(a) Let X be the roll of a fair die and $Y = \mathbf{1}\{X \geq 5\}$. What is $H(X \mid Y)$?

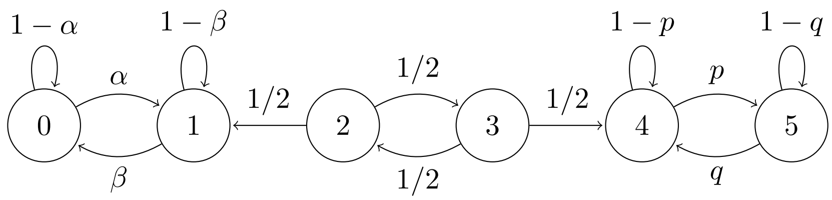
(b) Suppose the channel is a noiseless binary channel, i.e. $X \in \{0,1\}$ and $Y = X$. Use the theorem to find C.

(c) Consider a binary erasure channel with probability of erasure p. Use the theorem to find C.

Hint: To find the optimal p_X, it is helpful to let $p_X(1) = P(X = 1) = \alpha$.

3. Reducible Markov Chain

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



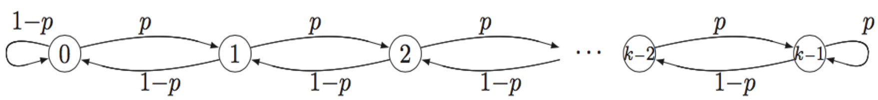
(a) Find all the recurrent and transient classes.

(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?

4. **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 $\rho = 1$.

5. **Metropolis-Hastings**

This problem proves properties of the **Metropolis-Hastings Algorithm**, which you saw in lab.

Recall that the goal of MH was to draw samples from a distribution $p(x)$. The algorithm assumes we can compute $p(x)$ up to a normalizing constant via $f(x)$, and that we have a proposal distribution $g(x, \cdot)$. The steps are:

- Propose the next state y according to the distribution $g(x, \cdot)$.
- Accept the proposal with probability

$$A(x, y) = \min\left\{1, \frac{f(y)}{f(x)} \frac{g(y, x)}{g(x, y)}\right\}.$$

- If the proposal is accepted, then move the chain to y; otherwise, stay at x.

(a) The key to showing why Metropolis-Hastings works is to look at the **detailed balance equations**. Suppose we have a finite irreducible Markov chain on a state space $\mathcal{X}$ with transition matrix P. Show that if there exists a distribution π on $\mathcal{X}$ such that for all $x, y \in \mathcal{X}$,

$$\pi(x)P(x, y) = \pi(y)P(y, x),$$

then π is a stationary distribution of the chain (i.e. $\pi P = \pi$).

(b) Now return to the Metropolis-Hastings chain. What is $P(x, y)$ in this case? For simplicity, assume $x \neq y$.

(c) Show $p(x)$, our target distribution, satisfies the detailed balance equations with $P(x, y)$, and therefore is the stationary distribution of the chain.

(d) If the chain is aperiodic, then the chain will converge to the stationary distribution. If the chain is not aperiodic, we can force it to be aperiodic by considering the **lazy chain**: on each transition, the chain decides not to move with probability $1/2$ (independently of the propose-accept step). Explain why the lazy chain is aperiodic, and explain why the stationary distribution is the same as before.

6. **(Optional) Relative Entropy and Stationary Distributions**

We define the *relative entropy*, also known as Kullback-Leibler divergence, between two distributions p and q as

$$D(p||q) = \mathbb{E}_{X\sim p}\left[\log\left(\frac{p(X)}{q(X)}\right)\right] = \sum_x p(x)\log\frac{p(x)}{q(x)}$$

(a) Show that $D(p||q) \geq 0$, with equality if and only if $p(x) = q(x)$ for all x. Thus, it is useful to think about $D(\cdot||\cdot)$ as a sort of distance function. *Hint:* For strictly concave functions f, Jensen's inequality states that $f(\mathbb{E}[Z]) \geq \mathbb{E}[f(Z)]$ with equality if and only if Z is constant.

(b) Show that for any irreducible Markov chain with stationary distribution π, any other stationary distribution μ must be equal to π. *Hint:* Consider $D(\pi||\mu P)$.