Name: Suresh Balakrishnama

Problem	Points	Score
1a	10	
1b	10	
1c	10	
1d	10	
2a	10	
2b	10	
3a	10	
3b	10	
3c	10	
3d	10	
Total	100	

Notes:

1. The exam is closed books/closed notes - except for one page (double-sided) of notes.

2. Please show ALL work. Answers with no supporting explanations or work will be given no credit.

3. Please indicate clearly your answer to the problem. If I can’t read it (and I am the judge of legibility), it is wrong. If I can’t follow your solution (and I get lost easily), it is wrong. All things being equal, neat and legible work will get the higher grade:)

Problem No. 1: Entropy Rate

(a) For the two-state Markov chain with the transition matrix, $P = \begin{bmatrix} 1-p_{01} & p_{01} \\ p_{10} & 1-p_{10} \end{bmatrix}$, find the entropy rate.

Solution:

The stationary distribution is represented by a vector μ whose components are the stationary probabilities of state 0 and state 1 as shown in Figure 1.

From Figure 1, we obtain:

$$\mu_0 p_{01} = \mu_1 p_{10}$$

Since $\mu_0 + \mu_1 = 1$, the stationary distribution is

$$\mu_0 = \frac{p_{10}}{p_{01}+p_{10}} \text{ and } \mu_1 = \frac{p_{01}}{p_{01}+p_{10}}$$

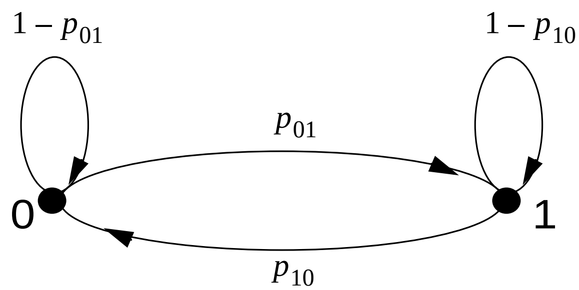


Figure 1: Two-state Markov chain

The entropy rate can be computed using the stationary distribution:

$$H\langle X_2|X_1\rangle = \mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{10}H(p_{01}) + p_{01}H(p_{10})}{p_{01}+p_{10}} bits$$

(b) Find the values of p_{01} and p_{10} that maximize the entropy rate.

Solution:

The entropy rate can be at most 1 bit since the process has only two states. To achieve this entropy rate the values should be

$$p_{01} = p_{10} = 0.5.$$

This process is a independent identically distributed process with

$$P(X_i = 0) = P(X_i = 1) = 0.5$$

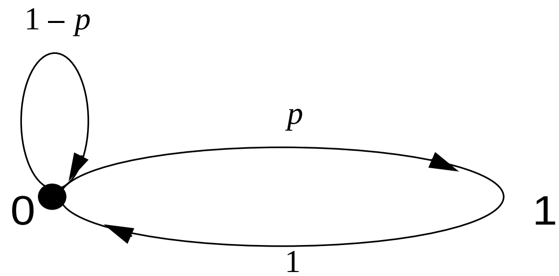


Figure 2: Two-state Markov chain

(c) For the two-state Markov chain with the transition matrix, $P = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix}$, find the maximum value of the entropy rate.

Solution:

The stationary distribution in this case is given by,

$$\mu_0 = \frac{p}{p+1} \quad \text{and} \quad \mu_1 = \frac{1}{p+1}$$

Using the formula for entropy rate,

$$H\langle X_2 | X_1 \rangle = \mu_0 H(p) + \mu_1 H(1) = \frac{H(p)}{p+1} bits$$

The maximum value of the entropy rate in this case is obtained by maximizing the value of p. The value of p should be less than 0.5, since state 0 allows more information to be generated than state 1. Hence, we can choose.

$$p = 0.4$$

Therefore, the maximum entropy rate is,

$$H(p) = 0.3776 bits$$

(d) Explain the reasons for any differences between the answers to (b) and (c), and any significance to this result. Your answer must show insight and understanding — don't simply tell me that the numbers are different because the probabilities are different.

Solution:

In case (b) the entropy rate is equally dependent upon both states 0 and 1. Both states are equiprobable. Entropy at both states contribute to the overall entropy. The transition from one state to another is significant. Since both states are equiprobable in this Markov chain the entropy value can obtained with a single state which is 1 bit.

In case (c) both states are not equiprobable and the entropy at each state accounts in different manner to the overall entropy. To explain more on this we can consider the transition among each state. The function tends to stay in the first state 0 for a considerable amount of period but at second state 1 the function immediately transits to the first state, thus the average weighting function is more dependent on the events at first state. This is the reason for the entropy in this case to be less than 1.

Problem No. 2: Data Compression

(a) Given an alphabet $S = \{S_1, S_2, \ldots, S_6\}$, and associated probabilities $P = \{p_1, p_2, \ldots, p_6\}$, you design an optimal D-ary code whose lengths turn out to be $L = \{1, 1, 2, 3, 2, 3\}$. Find a lower bound on D.

Solution:

We know that uniquely decodable codes satisfy Kraft's inequality. Therefore.

$\sum_i D^{-l_i} \le 1$ where l_i is specific word length.

For the problem given,

$$f(D) = D^{-1} + D^{-1} + D^{-2} + D^{-3} + D^{-2} + D^{-3} \le 1$$

Substituting different values for D, we get:

for $D = 2$, $f(D) = 7/4 > 1$ Kraft's inequality is not satisfied.

for $D = 3$, $f(D) = 26/27 < 1$ Kraft's inequality is satisfied.

So possible value of $D = 3$ since this value of D satisfies Kraft's inequality. Hence, we can choose this as lower bound.

(b) Let X be a binary-valued random variable with probabilities {ε,1-ε}. Estimate the expected length of an optimal code and discuss how this relates to the source coding theorem.

Solution:

The expected length of the optimal code should be $E(L) = \varepsilon \cdot 1 + (1 - \varepsilon) \cdot 1 = 1$.

Also we have $0 \le (H(X) = -\varepsilon \cdot \log\varepsilon - ((1-\varepsilon) \cdot \log(1-\varepsilon))) \le 1$,

with $H(X) = 1$ at $\varepsilon = \frac{1}{2}$

and $H(X) = 0$ at $\varepsilon = 0, 1$.

Based on the source coding theorem

$H(X) \le L^* < H(X) + 1$, when $\varepsilon = \frac{1}{2}$, we have $1 \le L^* < 2$.

where L^* is the length of optimal code.

Therefore, the expected length of 1 bit is the best result we can obtain for this optimal code. As a cross-check, when $\varepsilon = 0, 1$, we have $0 \le L^* < 1$, the expected length happens to be always less than 1, which is not the optimal case.

Problem No. 3: Gambling and Complexity

Consider a three-horse race with probabilities $(p_1, p_2, p_3) = (1/2, 1/4, 1/4)$. The odds associated with this race are $(r_1, r_2, r_3) = (1/4, 1/4, 1/2)$ (recall, these are typically set by how people bet on the race, not what the underlying odds are). The race is run many times, and number of the winning horse is transmitted over a communications channel using an optimal compression scheme.

(a) How many bits are required to transmit this information?

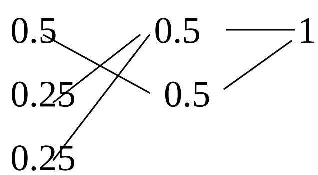
Solution:

Entropy in this case is,

$$H(p) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = 1.5 bits$$

Using Huffman code,

Codeword	
00	p_1
01	p_2
11	p_3



Expected Length is,

$$EL = 1(0.5) + 2(0.25) + 2(0.25) = 1.5 bits$$

Hence, $1.5 bits$ are required to transmit this information.

(b) Define an optimal betting strategy, (b_1, b_2, b_3), so that your wealth will increase as quickly as possible.

Solution:
Wealth will grow to infinity for the sets of (b_1, b_2, b_3) if and only if $W(b,p) > 0$ where,

$$W(b, p) = \left(\sum_i p_i \cdot \log(b_i \cdot o_i)\right)$$

$$= \sum_i p_i \cdot \log\left(\frac{b_i}{p_i} \cdot p_i \cdot \frac{1}{r_i}\right)$$

$$= \sum_i p_i \cdot \log\frac{1}{r_i} - H(p) - D(p \| b)$$

$$= \frac{7}{4} - 1.5 - D(p \| b)$$

$$= \frac{1}{4} - D(p \| b)$$

Based on the proportional gambling scheme, the wealth increases to infinity iff $b = p$ and the wealth in this case will be $2^{nW(b,\,p)} = 2^{n(0.25)}$.

(c) Is there only one solution to this problem? Explain.

Solution:
There can be only one solution for the wealth to increase to infinity since we know the true distributions. However, there can be several other solutions based on the betting strategy wherein the condition of $b = p$ doesn't hold good and the wealth rate can be computed in this case satisfying the criterion:

$$D(p \| b) < \frac{1}{4}$$

(d) Two three-horse races with different probabilities and odds are run, and the results are again reported over the communications link. For example, we transmit a pair of numbers, (a, b), indicating that the first race was won by horse a, and the second by horse b. Hence, you observe data such as $\{(1, 1), (2, 3), (1, 3), \ldots\}$. Estimate the Kolmogorov complexity of this sequence of numbers in terms of the Kolmogorov complexity of each race reported individually over the same link. In other words, how does the two-event process compare to two one-event processes?

Solution:

This is a problem of “complexity of sum”.

Kolmogorov complexity is given by:

$$K(x) = min l(p)$$

Kolmogorov complexity for each race is given by:

$$K(a) \le K\langle a|b\rangle + 2\log a + C$$

where $K\langle a|b\rangle$ is the conditional complexity of horse a, $2\log a$ represents the program length used to describe race of horse a and the digit 2 since each bit of $\log a$ is repeated to describe horse a and c is a constant.

Similarly, we can get the Kolmogorov complexity for race b

$$K(b) \le K\langle b|a\rangle + 2\log b + C$$

Overall complexity can be obtained by:

$$K(a+b) \le K(a) + K(b) + C$$

When transmitting programs for both race, we can write the output on the work tape of Turing machine, rather than on the output tape. Then, an instruction can be added to add the two lengths together and print them out on the output tape, thus minimizing the transmission length.