Name: Yu Zeng

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.

Let μ be the stationary distribution,

then from $\mu P = \mu$ we get: $\mu = \left(\frac{p_{10}}{p_{01}+p_{10}}, \frac{p_{01}}{p_{02}+p_{10}}\right)$

the entropy rate is:

$$H(\chi) = H(X_2|X_1) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij}$$

$$= \frac{1}{p_{01}+p_{10}}[p_{01}H(p_{10}) + p_{10}H(p_{01})]$$

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

The entropy rate is at most 1 bit because the process has only two states. The rate can be achieved if and only if $p_{01} = p_{10} = \frac{1}{2}$, in which case the process is actually i.i.d. with $Pr(X_i{=}0) = Pr(X_i{=}1) = \frac{1}{2}$.

(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.

From $\mu P = \mu$, we get: $\mu = \left(\frac{1}{1+p}, \frac{p}{1+p}\right)$.

$$H(\chi) = H(X_2|X_1) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij}$$

$$= \frac{1}{1+p} H(p)$$

to maximize $H(\chi)$, we take $\frac{\partial}{\partial p} H(\chi) = 0$, then get:

while $p = \frac{3-\sqrt{5}}{2}$ $\quad 1-p = \frac{\sqrt{5}-1}{2}$, $\quad H(\chi)_{max} = 0.693 \quad bits$

(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.

In problem 1(b), for each state, the next state can be either current state or the other state, depends on the probability.

But in problem 1(c), from state 2, it will not come back to state1, but just loop in the same state. This is not a random event, but a certain event, which reduce the overall entropy. So, the entropy rate from 1(c) is less than the one from 1(b).

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.

Since it's an optimal D-ary code, according to Kraft inequality, $\sum_i D^{-l_i} \le 1$ for this code, we have: $D^{-1} + D^{-1} + D^{-2} + D^{-3} + D^{-2} + D^{-3} \le 1$.
so, $D \ge 3$.

Example of code:

Table 1:

Probability	p1	p2	p3	p4	p5	p6
Codeword	0	1	20	220	21	221

(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.

For a binary-valued random variable X with probability {ε,1-ε}, The entropy is:

$H(X) = \varepsilon \cdot \log\frac{1}{\varepsilon} + (1-\varepsilon) \cdot \log\frac{1}{1-\varepsilon}$.

And the expected length of an optimal code is: $H(X) \le L \le H(X) + 1$.

So, $\varepsilon \cdot \log\frac{1}{\varepsilon} + (1-\varepsilon) \cdot \log\frac{1}{1-\varepsilon} \le L \le \varepsilon \cdot \log\frac{1}{\varepsilon} + (1-\varepsilon) \cdot \log\frac{1}{1-\varepsilon} + 1$

From this inequality, we will see that: when $\varepsilon \to 0$ or $\varepsilon \to 1$, $H(X) \to 0$, and $L \to 1$. These two case still satisfy above inequality, although the difference between L and $H(X)$ is large.

We can argue them as: for $\varepsilon \to 0$ or $\varepsilon \to 1$, these are two "almost" certain events. We almost don't need to see the code, but intuitively know the answer. So the entropy of these two case is so low. But, for a binary source, we still need 1 bit to distinguish ε and 1-ε. Actually 1 bit is the lower boundary of code length.

For the case ε = 1-ε = 1/2, $H(X)$ reaches its higher bound: 1 bit, and the expected length L is still 1 bit, which reaches the lower boundary of above inequality.

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?

Since the probabilities of winning horse is $(p_1, p_2, p_3) = (1/2, 1/4, 1/4)$, using Hoffman coding (which we have proved in class to be an optimal code), we get:

$$\frac{1}{2} \to 0$$

$$\frac{1}{4} \to 10$$

$$\frac{1}{4} \to 11$$

So, the expected length of codeword is: $EL = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{4} \times 2 = 1.5 \qquad bits$

2 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.

To make the wealth $S_n = 2^{nW(b,\,p)}$ increases, we need to let $W(b, p) > 0$.

Since $W(b, p) = E(\log S(X)) = \sum_i p_i \log b_i o_i = D(p||r) - D(p||b)$

And, $D(p||r) = \frac{1}{2} \cdot \log\frac{1/2}{1/4} + \frac{1}{4} \cdot \log\frac{1/4}{1/4} + \frac{1}{4} \cdot \log\frac{1/4}{1/4} = \frac{1}{4}$

to satisfy $W(b, p) = D(p||r) - D(p||b) = \frac{1}{4} - D(p||b) > 0$, we just need to restrict $D(p||b)$ as $D(p||b) < \frac{1}{4}$.

While $D(p||b) = 0$, $W(b, p)_{max} = \frac{1}{4}$, so $S_n = 2^{nW(b,\,p)}$ will increase as quickly as possible. i.e. $p = b$. Or, you can say, betting according to the proportional distribution is the optimal strategy.

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

The doubling rate is $W(b, p) = D(p||r) - D(p||b)$, which means it is the difference between the distance of the bookie's estimate from the true distribution and the distance of the gambler's estimate from the true distribution. Hence the gambler can make money only if his estimate b is better than the bookie's. So, for this problem, if $D(p||b) < \frac{1}{4}$, your wealth will increase. Your strategy will always be minimize the distance between p and b.

(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 $\{(1, 1), (2, 3), (1, 3), \dots\}$. 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?

Suppose the program to transmit a is p_a, the program to transmit b is p_b, with Kolmogorov complexity $K(a)$ and $K(b)$ respectively. To transmit event (a, b), we use program p_{a+b}, and its complexity is $K(a, b)$. Then this problem is equal to compare $K(a, b)$ and $K(a) + K(b)$.

The program p_{a+b} can be: run the program p_{a+b} to transmit (a, b) and then halt.

The program of two-event process can be: run program p_a to transmit a, and then run program p_b to transmit b, then halt.

Hence, we see that the Kolmogorov complexity of two one-event processes is greater than the two-event process. i.e.: $K(a+b) \le K(a) + K(b) + c$.