Name: ZHAO JIE

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:

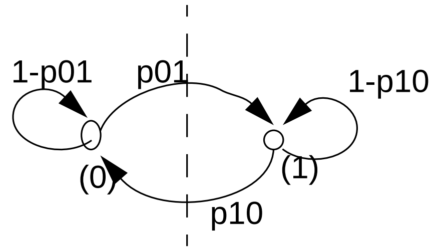
1. The exam is closed books/closed notes - except for one page (double-sided) of notes.

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.



Calculate the stationary probability,

$$u_0 + u_1 = 1$$

$$u_0 \cdot p_{01} = u_1 \cdot p_{10}$$

so $\mu_0 = \dfrac{p_{10}}{p_{10}+p_{01}}$, $\mu_1 = \dfrac{p_{01}}{p_{10}+p_{01}}$

For Markov Chain, entropy rate is:

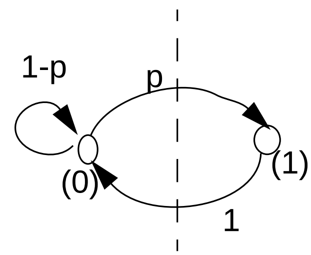
$$H(\chi) = H(x_2|x_1) = \mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{10}}{p_{10}+p_{01}}H(p_{01}) + \frac{p_{01}}{p_{10}+p_{01}}H(p_{10})$$

(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. Since the effects of p_{01} and p_{10} are same, the maximum value of $H(\chi)$ must be reached when $p_{01} = p_{10} = \frac{1}{2}$. So the maximum value of entropy rate is,

$$H(\chi) = \frac{1}{2}\times 1 + \frac{1}{2}\times 1 = 1$$

(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.



In this case, $\mu_0 = \frac{1}{1+p}$, $\mu_1 = \frac{p}{1+p}$

Entropy rate is

$$H(\chi) = \mu_0 H(p) + \mu_1 H(1) = \frac{H(p)}{1+p} = \frac{1}{1+p}\left((1-p)\log\frac{1}{(1-p)} + p\log\frac{1}{p}\right),$$

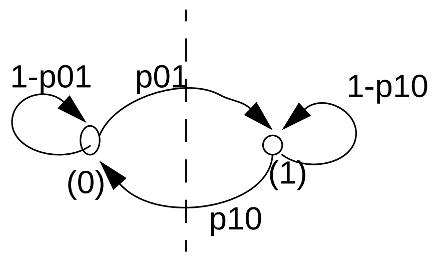
differentiate $H(\chi)$, the maximum value is reached when $\frac{dH(\chi)}{dp} = 0$

we find the maximum value of $H(\chi)$ occurs for $p = \frac{3-\sqrt{5}}{2} = 0.382$, then

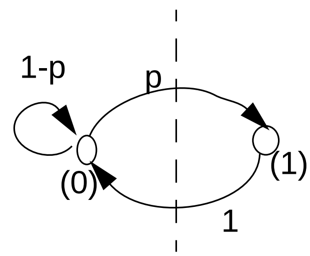
So the maximum value of entropy rate is:

$$H(\chi) = \frac{H(p)}{1+p} = \frac{1}{1.382} \cdot H(0.382) = \frac{0.958}{1.382} = 0.694.$$

(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.



(b)



(c)

In (b), the maximum entrpy rate is 1 and is achieved when $p_{01} = p_{10} = \frac{1}{2}$.

In (c), if X is at state 1, at the next step, it must transit to state 0, thus it is a certain event, the conditional entorpy on state 1 is 0. Hence the entropy will decrease with time, the maximum entropy rate should be less than 1. In addition, because the 0 state permits more information to be generated than 1 state, it is resonable that the probability p which achieves the maximum entropy rate should be less than 1/2.

Problem No. 2: Data Compression

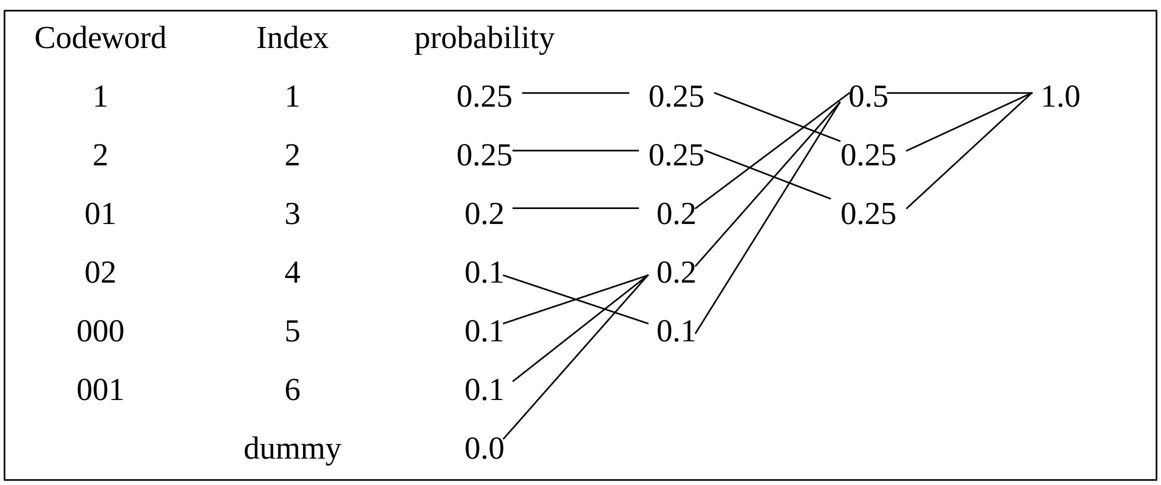
(a) Given an alphabet $S = \{S_1, S_2, \ldots, S_6\}$, and associated probabilities $P = \{p_1, p_2, \ldots, p_6\}$, you design an optimal D-ary code whose lengths turn out to be $L = \{1, 1, 2, 3, 2, 3\}$. Find a lower bound on D.

According to Kraft inequality, $\sum D^{-li} \le 1$,

when D=2, $\sum D^{-li} = 2 \cdot D^{-1} + 2 \cdot D^{-2} + 2 \cdot D^{-3} = 1.75 > 1$

when D=3, $\sum D^{-li} = 0.963 < 1$, Satisfy Kraft inequality, so the lower bound on D is 3.

For example, design a Huffman code with length $L = \{1, 1, 2, 3, 2, 3\}$,



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

The expected length of an optimal code $H(x) \le L < (H(x) + 1)$,

so for a binary-valued random variable X with probabilities {1,1-ε},

Entropy is $H(x) = \varepsilon \cdot \log\frac{1}{\varepsilon} + (1-\varepsilon) \cdot \log(\frac{1}{1-\varepsilon})$

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

We can construct an optimal code: let 0 for $p = 1$ and 1 for $p = 1-\varepsilon$.So the codeword length is $L = \sum p_i \cdot l_i = \varepsilon + 1 - \varepsilon = 1$.

When $\varepsilon \to 0$, the probability of X =1 (or X=0) is close to 1. The entropy of X $H(x) \to 0$.

But the length of the optimal code is 1 bit, which is almost 1 bit above its entropy, so it approaches the upper bound.

Whenf $\varepsilon = \frac{1}{2}$, $H(x) = 1$. the length of the optimal codeword we constructed is 1, that is equal to its entropy, so it reaches the lower bound.

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?

$$H(x) = \sum_i p_i \cdot \log\frac{1}{p_i} = \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1.5$$

The expected code word length must be greater or equal to the entropy, and must be integer, so 2 bits are required.

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

The gambler can make money only when the doubling rate $W(b, p) > 0$,

$$W(b, p) = E(\log S(X)) = \sum_{k=1}^{m} p_k \cdot \log(\mathrm{b}_k \cdot o_k)$$

$$= \sum_{k=1}^{m} p_k \cdot \log\left(\frac{b_k}{p_k}\cdot\frac{p_k}{r_k}\right) = D(p//r) - D(p//b) > 0$$

$$D(p//\mathrm{r}) = \frac{1}{2}\cdot\log\left(\frac{1/2}{1/4}\right) + \frac{1}{4}\cdot\log\left(\frac{1/4}{1/4}\right) + \frac{1}{4}\cdot\log\left(\frac{1/4}{1/2}\right) = \frac{1}{4}$$

So, it needs relative entropy $D(p//\mathrm{b}) < \frac{1}{4}$ in order to increase the wealth.

When $D(p//\mathrm{b}) = 0$, the $W(b, p)$ reaches the maximum, the wealth increases most quickly. So the best betting strategy is $b = p$, i.e. $b_1 = \frac{1}{2}$, $b_2 = \frac{1}{4}$ and $b_3 = \frac{1}{4}$. This shows that proportional gambling is optimal.

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

If this “problem” refers to the strategy that can make the wealth increase, the answer is NO, because as long as the strategy can make $D(p//\mathrm{b}) < \frac{1}{4}$, the wealth will increase.

If this “problem” refers to the best strategy that makes the wealth increase most quickly, the answer is YES, because only proportional gambling strategy can make $D(p//\mathrm{b}) = 0$.

(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?

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

Let the Kolmogorov complexity of the numbers a and b are $K(a)$ and $K(b)$ respectively. Then there must exist two programs P_x and P_y, of length $K(a)$ and $K(b)$ respectively, which can transmit a and b. Form the following program:

Run the following two programs, not halting after the first; Run the program P_x to transmit a, interpreting the halt as a jump to the next step; Run the program P_y to transmit b.

This program with length $K(a) + K(b) + C$, transmits the numbers (a, b). Hence

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