Name:

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.

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

so the entropy rate is:

$$\mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{01}}{p_{10}+p_{01}} H(p_{01}) + \frac{p_{10}}{p_{01}+p_{10}} H(p_{10})$$
$$= \frac{p_{01}}{p_{10}+p_{01}}\left(p_{01}\log\frac{1}{p_{01}} + (1-p_{01})\log\frac{1}{1-p_{01}}\right) + \frac{p_{10}}{p_{10}+p_{01}}\left(p_{10}\log\frac{1}{p_{10}} + (1-p_{10})\log\frac{1}{1-p_{10}}\right)$$

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

From (a), we know entropy rate equal to

$$\mu_0 H(p_{01}) + \mu_1 H(p_{10}) = \frac{p_{01}}{p_{10}+p_{01}}\left(p_{01}\log\frac{1}{p_{01}} + (1-p_{01})\log\frac{1}{(1-p_{01})}\right) + \quad ,$$
$$\frac{p_{10}}{p_{10}+p_{01}}\left(p_{10}\log\frac{1}{p_{10}} + (1-p_{10})\log\frac{1}{(1-p_{10})}\right)$$

when $p_{01} = \frac{1}{2}$ and $p_{10} = \frac{1}{2}$, the entropy rate get the maximum value,

$$\frac{1}{2}\times 1 + \frac{1}{2}\times 1 = 1$$

(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.

Still using $H(\chi) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij} = \sum_i \mu_i H(x_i)$

here, since H(1,0)=0, so entropy rate $H(\chi) = \frac{1}{1+p}\left((1-p)\log\frac{1}{(1-p)} + p\log\frac{1}{p}\right)$, to get the

maximum value of entropy rate, we should make $\frac{dH(\chi)}{dp} = 0$ and $\frac{d^2H(\chi)}{dp^2} > 0$

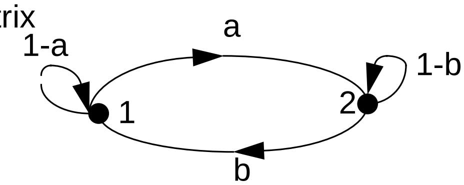
solve these two equation, we find $p = \frac{3-\sqrt{5}}{2}$, the maximum value of entropy

rate is: 0.50

(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.

The figure on the right shows the stationary two state Markov chain with a transition matrix

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$$



In (c), b=1, which means this is a certain event, entropy for this event is 0. So the entropy of the whole process decreases.

Since entropy rate means per symbol entropy of the n random variables, it decrease accordingly. In (c), since there is no information in state 2, the maximum entropy rate drop to 1/2 of the maximum entropy rate in (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.

From Kraft inequality, we know for a prefix code, $\sum D^{-li} \le 1$, when D=2, $\sum D^{-li} = 1.75 > 1$ while when D=3, $\sum D^{-li} = 0.96 < 1$, so the lower bound on D is 3.
We can use Huffman code since it's optimal. The code are shown in the figure and Table.

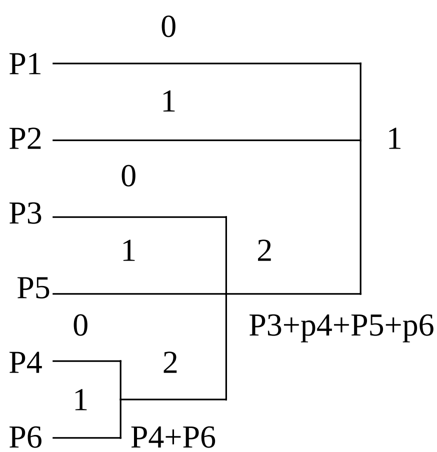


Table 1:

Probability	Codewords
P1	0
P2	1
P3	20
P4	220
P5	21
P6	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 an optimal code, the expected length $H(x) \le L < (H(x) + 1)$, so for a binary-valued random variable with probabilities {1,1-ε}, Entropy is $H(x) = -(\varepsilon\log_2\varepsilon + (1-\varepsilon)\log_2(1-\varepsilon))$
the expected length is between $-(\varepsilon\log_2\varepsilon + (1-\varepsilon)\log_2(1-\varepsilon))$ and $1-(\varepsilon\log_2\varepsilon + (1-\varepsilon)\log_2(1-\varepsilon))$

In source coding, we want to minimize the expected length, so we want to code length to be near or equal to entropy.

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?

We know that Huffman code are optimal, we can use it to transmit this information,

$$H(x) = -\sum_i P_i \log P_i = -\left(\frac{1}{2}\times\log\left(\frac{1}{2}\right)+\frac{1}{4}\times\log\left(\frac{1}{4}\right)+\frac{1}{4}\times\log\left(\frac{1}{4}\right)\right) = 1.5$$

since $H_D(x) \le L^* < H_D(x) + 1$,

so 2 bits are needed.

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

To make you wealth increase, we have to make doubling rate $W(b, p) > 0$,

i.e. $W(p, b) = E(\log(s(x))) = \sum_{k=1}^{m} p_k \log(\mathrm{b}_k O_k) = D(p//r) - D(p//b) > 0$

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

so, $D(p//\mathrm{b}) < \frac{1}{4}$ is required to increase your wealth

To get the wealth increase as quickly as possible, we need to get the optimum doubling rate. It's proved that when $b^* = P$, i.e. $D(p//\mathrm{b}) = 0$, (which means the gambler bets on each horse in proportion to it's probability of winning, the doubling rate is optimal, so the best strategy here is $b_1 = 1/2$, $b_2 = 1/4$ and $b_3 = 1/4$

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

There are more than one solutions to win. As long as $D(p//\mathrm{b}) < \frac{1}{4}$, which means $W(b, p) > 0$, you will grow your wealth. But to make your wealth grow fast, there is only one solution, you need to choose b =p, so that the $D(p//\mathrm{b}) = 0$ and $W(b, p)$ gets the maximum value.

(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?

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

Suppose the Kolmogorov complexity of the program P_x to transmit a is $K(a)$, and that of program P_y which transmit b is $K(b)$,

we can run these two program like this: Run the program P_x to transmit a, interpreting the halt as a jump to the next stop, run the program P_y to transmit b until halt.

So the relationship between the kolmogorov complexity of the two-event process and those of two one-event processes is: $K(a, b) \le K(a) + K(b) + C$