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.

Assume the states to be $\vec{\mu} = [\mu_0, \mu_1]^T$

then we have $\mu_0(1-p_{01}) + \mu_1 \cdot p_{10} = \mu_0$ and $\mu_1(1-p_{10}) + \mu_0 \cdot p_{01} = \mu_1$

Solve these equations we get $\mu_0 = \dfrac{p_{10}}{p_{01}+p_{10}}$, $\mu_1 = \dfrac{p_{01}}{p_{01}+p_{10}}$

Hence the entropy rate is $H(\chi) = \mu_0 \cdot H(p_{01}) + \mu_1 \cdot H(p_{10})$

(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. We can achieve this rate when $p_{01} = \dfrac{1}{2}$ and $p_{10} = \dfrac{1}{2}$. That is,

$$H(\chi) = \mu_0 \cdot H(p_{01}) + \mu_1 \cdot H(p_{10}) = \mu_0 \cdot H\left(\frac{1}{2}\right) + \mu_1 \cdot H\left(\frac{1}{2}\right) = \mu_0 + \mu_1 = 1$$

(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.

According to (a), we have

$$H(\chi) = \mu_0 \cdot H(p_{01}) + \mu_1 \cdot H(p_{10}) = \mu_0 \cdot H(p) + \mu_1 \cdot H(1)$$
$$= \mu_0 \cdot H(p) = \frac{H(p)}{p+1}$$

By calculus, we find that at $p = \frac{(3-\sqrt{5})}{2} = 0.382$, $\frac{\partial}{\partial p} H(\chi) = 0$ and $\frac{\partial^2}{\partial p^2} H(\chi) < 0$.

Therefore, we get the maximum value of the entropy rate as $H(\chi) = \frac{H(p)}{p+1} = 0.694 bits$

(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.

The maximum value of the entropy rate in (c) is less than that in (b). This is because in (c), state 1 will never remain itself - it will go back to state 0 with a probability of 1. As a result, the uncertainty in this case is reduced, hence the maximum value of the entropy rate is decreased.

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.

An optimal code should satisfy the Kraft inequality $\sum_i D^{-l_i} \le 1$.

For $D = 2$, $\sum_i 2^{-l_i} = 1.75 > 1$, the Kraft inequality doesn't hold.

For $D = 3$, $\sum_i 3^{-l_i} = 0.96 < 1$, the Kraft inequality holds.

Therefore the lower bound on D is $D = 3$.

(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.

Since X is a binary-valued random variable, the optimal code could be

$(X = x_0) \to 0, (X = x_1) \to 1$

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

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 $\varepsilon = 0, 1$.

According to the source coding theorem $H(X) \le L^* < H(X) + 1$, when $\varepsilon = \frac{1}{2}$, we have $1 \le L^* < 2$, hence the expected length of 1 we got for the optimal code is the best result we can get; while when $\varepsilon = 0, 1$, we have $0 \le L^* < 1$, hence the result we got stands for the worst 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?

$$H(X) = -\left(\frac{1}{2}\cdot\log\frac{1}{2}\right)-\left(\frac{1}{4}\cdot\log\frac{1}{4}\right)-\left(\frac{1}{4}\cdot\log\frac{1}{4}\right) = 1.5bits$$

According to the source coding theorem, this is the least number of bits required to transmit this information.

We can build the following Huffman code to transmit the information,

$horse1 \rightarrow 0$, $horse2 \rightarrow 10$, $horse3 \rightarrow 11$

and the expected length of the code is exactly $L = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 1.5bits$

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

The doubling rate is

$$W(b, p) = \sum_i p_i \cdot \log(b_i \cdot o_i) = \sum_i p_i \cdot \log\left(\frac{b_i}{p_i}\cdot p_i \cdot \frac{1}{r_i}\right)$$

$$= \sum_i p_i \cdot \log\frac{1}{r_i} - H(p) - D(p \| b) = \frac{7}{4} - 1.5 - D(p \| b) = \frac{1}{4} - D(p \| b)$$

For wealth to increase, the doubling rate should be greater than 0, that is, $\frac{1}{4} - D(p \| b) > 0$, or $D(p \| b) < \frac{1}{4}$. Any betting strategy (b_1, b_2, b_3), if satisfies this inequality, will result in an increasing wealth.

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

No. There could be many betting strategies with which the wealth will increase, if only $D(p \| b) < \frac{1}{4}$ holds.

(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?

Assume the Kolmogorov complexity to be $K((a, b))$, $K(a)$, and $K(b)$ respectively for the three processes.

Use the following program to print $\{(a, b)\}$: run program I, printing result of race I; program I halt; call program II, printing result of race II; program II halt; return to program I, continue to print the next pair of result.

In this program, we only have some extra instructions to call and return from program II. Therefore, we have the following relationship $K((a, b)) = K(a) + K(b) + c$, where c is a constant.