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:

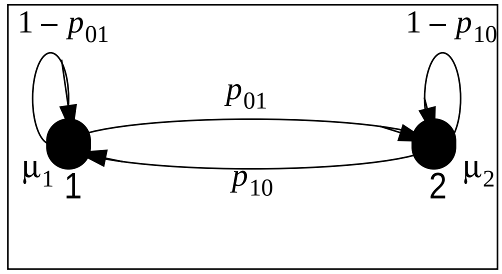
1. The exam is closed books/closed notes - except for one page (double-sided) of notes.

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.



The adjacent figure represents the above Markov chain. We first calculate the stationary distribution μ.

We know that $\mu P = \mu$ For a 2 state markov chain, this is the same as $\mu_1 p_{01} = \mu_2 p_{10}$. Also $\mu_1 + \mu_2 = 1$.

Solving yields,

$\mu_1 = p_{10}/(p_{10}+p_{01})$ and $\mu_2 = p_{01}/(p_{10}+p_{01})$

The entropy rate for a stationary markov chain is $H(\chi) = -\sum_i \mu_i \sum_j p_{ij} \log p_{ij}$

So, for our example, the entropy rate is $H(\chi) = \mu_1 H(p_{01}) + \mu_2 H(p_{10})$, where $H(p) = -(p\log p + (1-p)\log(1-p))$.

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

Since $H(\chi)$ is symmetric with respect to p_{10} and p_{01}, it's maximum value is reached when $p_{10} = p_{01}$. Let $p_{10} = p_{10} = p$. This also results in $\mu_1 = \mu_2 = 0.5$. The equation for $H(\chi)$ derived in the precious section also suggests that $H(\chi)$ will be a maximum if $H(p_{01}) = H(p_{10}) = H(p)$ is a maximum. Now, $H(p) = -(p\log p + (1-p)\log(1-p))$ is maximum for p=0.5. So $p_{01} = p_{10} = 0.5$. Substituting this and estimating $H(\chi)$ yields $H(\chi) = H(p) = -(0.5\log 0.5 + (1-0.5)\log(1-0.5)) = 1bit$.

(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.

Similar to the argument for the part a, we first estimate the stationary distribution. For this problem, we can write

$\mu_1 p = \mu_2 1$. Also we know that $\mu_1 + \mu_2 = 1$. So we have $\mu_1 = \frac{1}{1+p}$

Now the entropy rate $H(\chi)$ can be written as

$H(\chi) = \mu_1 H(p) + \mu_2 H(1)$, where $H(p) = -(p \log p + (1-p)) \log(1-p)$ and

$H(1) = -(0 \log 0 + 1 \log 1) = 0$. So, simplifying and rewriting the equation for $H(\chi)$, we have

$H(\chi) = \frac{-1}{1+p}(p \log p + (1-p) \log(1-p))$. The maximum value of $H(\chi)$ will be obtained by setting the derivative of this equation with respect to p equal to zero and substituting the obtained value of p. Setting the derivative equal to zero yields $p = 0.38$. Substituting this value for $H(\chi)$ yields $max H(\chi) = 0.4$.

(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 major difference is the stationary distribution. In part b, it was a uniform distribution. So the starting state could have been 1 or 2 with equal probability. In part c, there is more chance of the starting state being state 2, which is a certain event. So the entropy rate, which can be visualized as the expected value of the entropy of the system, will be less as $H_{part(c)} < H_{part(b)}$. In other words, the entropy rate in the second case will be less as one of the components in the formula for this is zero.

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 the Kraft inequality, we know that $\sum_i D^{-l_i} \le 1$, where l_i are the individual codeword lengths. For this case it turns out to be $D^{-1} + D^{-1} + D^{-2} + D^{-3} + D^{-2} + D^{-3} \le 1$, or $D^{-1} + D^{-2} + D^{-3} \le 1$. Substituting $D = 2$, does not satisfy this inequality. $D = 3$ satisfies this inequality.

Hence $D = 3$ is a lower bound on 3.

A 3-ary code whose lengths are consistent with the ones given in the question is given below:

Probability	p_1	p_2	p_3	p_4	p_5	p_6
3-ary 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.

Using Kraft's inequality $\sum D^{-l_i} \le 1$ and trying to minimize the expected codeword length $\sum p_i l_i$, we obtain the optimal codeword lengths as $l_i^* = -\log_D p_i$. So for our example we have $l_1^* = -\log_2 \varepsilon$ and $l_2^* = -\log_2(1-\varepsilon)$. Also the expected codeword length is $L^* = \sum p_i l_i^* = H_D(X)$. It can be noted that as $\varepsilon \to 0$ or $\varepsilon \to 1$, then $H(X) \to 0$ and $L^* \to 1$.

The source coding theorem states that the optimal code length lies between $H_D(X)$ and $H_D(X) + 1$. The lower bound has already been proved above. The upper bound is because the values of l_i^* are constrained to be integers. The proof for the upper bound is also standard.

The values we had obtained for l_1^* and l_2^* are also consistent with the source coding theorem.

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?

The minimum number of bits required to transmit this information can be estimated by calculating the entropy of the distribution. This value of the entropy is also the minimum number of bits required for transmission.

Hence, minimum number of bits required = $H(p) = -(0.5\log 0.5 + 0.25\log 0.25 + 0.25\log 0.25)$
$= 1.5$ bits/race.

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

The most optimal betting strategy would be to follow the proportional betting strategy. Here we make $(b_1, b_2, b_3) = (p_1, p_2, p_3)$.

The reason that the proportional betting strategy is the most optimal is due to the fact that the doubling rate $W(b, p)$ is given by $W(b, p) = \sum_{k=1}^{m} p_k \log b_k o_k \le \sum p_i \log o_i - H(p)$. The equality holds only if ***p* = *b***. Thus the doubling rate is a maximum for proportional betting and hence this strategy is the most optimal.

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

There might be other strategies to increase our wealth but the proportional betting strategy is the most optimal one. The basis on which we can make this statement is the fact that the doubling rate can also be expressed as $W(b, p) = D(p \| r) - D(p \| b)$. In other words, the doubling rate is the difference between the bookies estimate from the true distribution and the distance of the gamblers estimate from the true distribution.
Hence the gambler can make money only if his estimate is better than the bookie's. Also if the gamblers estimate is equal to the true distribution, he gains the most, which means that the proportional betting strategy is the most optimal.

(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?

Let $K(a)$ be defined as the Kolmogorov complexity of the number of times horse "*a*" wins the race. Similarly let us define $K(b)$.

Now there will surely exist programs p_a and p_b that will print events *a* and *b*. Based on this, let us define the program

"Execute p_a till a halt is reached and then jump to p_b and execute it.".

This is the event (*a*,*b*) and it's Kolmogorov complexity is $K(a, b) \le K(a) + K(b) + c$, where c is a constant. This equation can help us compare the two event process to the two one event processes.
The complexity can also be related to the compression of the transmitted data. The complexity would be less in case the data was transmitted jointly. This is true as more compression can be achieved based on the fact that $H(X, Y) \le H(X) + H(Y)$.