Name: Richard Duncan

Problem	Points	Score
1a	10	
1b	10	
1c	10	
2a	10	
2b	10	
2c	10	
2d	10	
3a	10	
3b	10	
3c	10	
Total	100	

Notes:

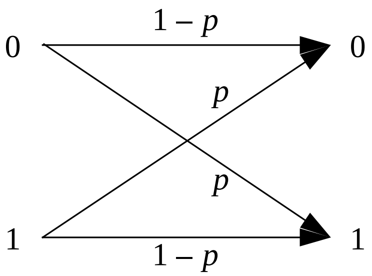
1. The exam is closed books/closed notes - except for one page (double-sided) of notes.

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: Channel and Source Coding

(a) For a binary symmetric channel, show that $I(X;Y) \le 1 - H(p)$.



Let X be the input symbol and Y be the output symbol. From the figure it is clear that p is the probability of error in the transmission. We define channel capacity to be $C = \max I(X;Y)$. The mutual information can be bounded

by $$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum p(x)H(Y|X = x) \\ &= H(Y) - \sum p(x)H(p) \\ &= H(Y) - H(p) \\ &\le 1 - H(p) \end{aligned}$$

Since only one bit is being transmitted, obviously $H(Y) \le 1bit$.

(b) Suppose the letters {a,b,c,d} are transmitted over this channel, and these letters have a prior distribution of $p(x) = \{1/4, 1/4, 1/8, 3/8\}$. Discuss the best way to send this data over the channel such that you minimize the error rate and you minimize the number of bits transmitted.

The entropy of the source $H(p(x)) = -\langle \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{8} + \frac{3}{8}\log\frac{3}{8} \rangle = 1.906bits$. One good approach is to consider the data compression and data transmission problems separately. The joint source channel coding theorem tells us that this two step approach is as good as any other method for transmitting the data over a noisy channel. Since we know the pdf of the source, we can use a Huffman code to compress the data rate to nearly the entropy. However, since this is a noisy channel we must add back redundancy to approach error free communication.

The channel capacity will still be bounded by $C = \max I(X;Y) = H(Y) - H(p)$. This is for each bit transmitted, though, and we now need more than one bit to discriminate between the output classes. So, the channel will cost us $H(p)l_x$, where l_x is the expected length of the optimal code. From data compression and competititive optimality we know that this length will be bounded by the entropy and entropy plus one. Capacity now becomes$C = H(Y) - l_x H(p)$.

(c) Suppose we cascade a second BSC with the same properties as the first. Derive an expression for the capacity, state whether the capacity increases or decreases, and explain why.

If we must transmit the data over two such BSCs, the capacity will be unchanged. Whatever redundancy is added to the code words to be nearly error free in the first channel will also work in the second channel, you need add no more redundancy. The probability of error, however, will double as you have two independent events. This error is near zero, though, and will have little bearing.

Problem No. 2: Continuous Random Variables

(a) Prove the scaling theorem for the entropy of a continuous random variable.

Theorem: $H(aX) = h(X) + \log|a|$

Proof: Let $Y = aX$, then $f_y(y) = \frac{1}{|a|} f_x\left(\frac{y}{a}\right)$. So,

$$h(aX) = -\int f_y(y) \cdot \log f_y(y) dy = -\int \frac{1}{|a|} f_x\left(\frac{y}{a}\right) \cdot \log\left(\frac{1}{|a|} f_x\left(\frac{y}{a}\right)\right) dy$$

$$= -\int f_x(x) \cdot \log f_x(x) + \log|a| = h(X) + \log|a|$$

(b) Derive an expression for the capacity of a power-limited Gaussian channel (hint: compute the mutual information in terms of the entropies of the signal and noise, and apply bounds for these entropies).

$C = \max I(X;Y), EX^2 \le P$

$$I(X;Y) = h(Y) - h(Y|X)$$
$$= h(Y) - h(X + Z|X)$$
$$= h(Y) - h(Z|X)$$
$$= h(Y) - h(Z)$$

For a Gaussian channel, $h(Z) = \frac{1}{2}\log 2\pi e N$.

We also need to expand the power constraint, $EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N$, since X and Z are independent and the zero mean noise has $EZ = 0$. Since $EY^2 = P + N$, $H(Y)$ is bounded by $H(Y) \le \frac{1}{2}\log 2\pi e(P+N)$ since a Gaussian is maximum entropy. If we apply this back to the capacity, $I(X;Y) = h(Y) - h(Z) \le \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$.

(c) Explain the significance of this result on three types of problems: compression, system identification, and maximum entropy spectral estimation.

If we wanted to get very fast error free communication, we need only spike up the power over the line to get great separation in the two classes. Unfortunately, we have a power constraint.

When trying to identify a system, we make the Gaussian assumption. In doing so we need only fix two parameters. If we assumed any other distribution we would need to assume more things about the distribution, hence the Gaussian assumption is on the average our best choice.

Spectral estimation is covered in 2d.

(d) Explain Burg's Maximum Entropy Theorem.

Burg's ME theorem was a fundamental breakthrough by relating the spectral estimation problem back to information theory. We already knew that the Gaussian model was best from the auto-correlation proofs, Burg proved this result may also be obtained through information theory as the maximum entropy process. His work opened up the door for us to use information theory to solve many problems: many HMM training algorithms use entropy now as an internal metric.

He stated that the entropy rate subject to the auto correlation constraints $R_0, R_1, \ldots, R_p$ is maximized by the p th order zero-mean Gauss-Markov $X_i = \sum_{k=1}^{p} a_k x_{i-k} + z_i'$. The ME spectrum is $S(l) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} \alpha_k e^{-ikl}\right|^2}$.

Maximum entropy works because it is the worst case. Any other distribution will be more predictable than what we assume. Also, if we assume a Gaussian we are placing the fewest number of constraints on the system. If we are guessing, it is best to guess as few parameters as possible. A Gaussian distribution is completely described by its mean and variance, any other type of distribution has more parameters.

Problem No. 3: Statistics

Consider a six-sided die containing the numbers {1,2,3,4,5,6}. You roll this die ten times and generate the sequence {1,2,3,4,5,6,2,4,6}.

(a) Describe the type class for this event.

The type class of P, denoted $T(P)$, is the set of sequences of length n and type $P \in P_n$.

The type class of P_x is the set of all sequences of length 9 with two instances of each even number and one instance of each odd number, one such type contained in this class is $P_x = \left(\frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}\right)$. So, if we subscribe to the theory that $P_x = Q$, the die seems to be twice as likely to produce even numbers as to produce odd numbers. Since n is so small, however, we could very easily have a fair die as well.

(b) Bound the size of the type class.

$$H(P_x) = 3\left(-\frac{1}{9}\log\frac{1}{9}\right) + 3\left(-\frac{2}{9}\log\frac{2}{9}\right) = 2.50 bits$$

The size of the type class may be bounded by

$$\frac{2^{nH(P_x)}}{(n+1)^{|X|}} \le |T(P)| \le 2^{nH(P_x)},$$

$$\frac{2^{9(2.5)}}{(9+1)^6} \le |T(P)| \le 2^{9(2.5)},$$

Therefore, $5.93 \le |T(P)| \le 5.93x10^6$

(c) Discuss the different ways to estimate the probability of the event above using concepts developed in this course. Be as precise as possible. Do not assume this is a fair die.

One way would be to use relative entropy. If we some how knew $D(P_x||Q)$ we could use the equation $Q^n(X) = 2^{n(-D(P_x||Q) + H(P_x))}$. From the law of large numbers we know that as $n \to \infty$ $D(P_n||Q) \to 0$ with probability one.

Of course, basic probabilities and counting techniques could be used, but this would quickly get cumbersome as the problem became more complicated. From a probability background we could also obtain confidence intervals (using a Chi-squared table) to show how likely such an event is with a fair die.