Name:

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 $I(X;Y) \le 1 - H(p)$**.**

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

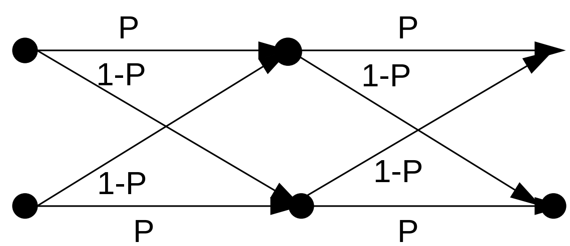
(b) Suppose the letters {a,b,c,d} are transmitted over this channel, and these letters have a prior distribution of $p(x) = \{1/4, 1/4, 1/8, 3/8\}$**. Discuss the best way to send this data over the channel such that you minimize the error rate and you minimize the number of bits transmitted.**

This is a question about source coding and channel coding. For source coding, we want to use the code with minimum expected length. Since we know the distribution of the letters, we can use Huffman coding so that we can use the minimum number of bits to transmit the information.

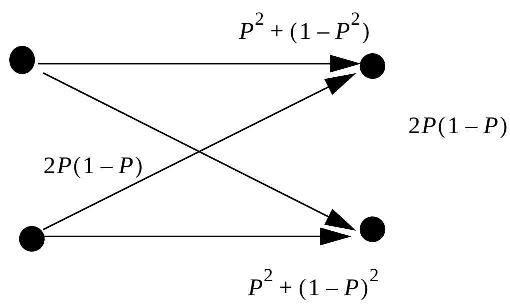
To achieve the lowest error rate, according to channel coding theorem, for every rate R<C (C is the channel capacity) , there exists a sequence $\langle 2^{nR}, n\rangle$ code with maximum probability of error $\lambda^{(n)} \to 0$, we may choose Hamming code to achieve a low error rate. In real application, we can choose joint source and channel coding.

(c) Suppose we add 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.

Two Binary Symmetric connected in cascade is given below:



This is equivalent to a single symmetric channel:



So the capacity of the cascade of two channel is:

$1 - H(P^2 + (1 - p)^2)$

The capacity of the cascade of two BSC channel decreases compared with that of the single BSC channel since when there are more cascaded channels, the chance of errors will increase and cause the mutual information between start and end decrease.

Problem No. 2: Continuous Random Variables

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

The theorem is: $h(aX) = h(X) + \log|a|$

Let $Y = aX$, Then $f_y(y) = \frac{1}{|a|} f_Y\left(\frac{y}{a}\right)$, and

$$\begin{aligned} h(aX) &= -\int f_Y(y)(\log f_Y(y))dy \\ &= -\int \frac{1}{|a|} f_X(x)\log f_X(x) + \log|a| \\ &= -\int f_x(x)\log f_x(x) + \log|a| \\ &= h(X) + \log|a| \end{aligned}$$

(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).

To the get power limited capacity, $C = max \quad I(X;Y)$, $EX^2 \le P$

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \end{aligned}$$

For a Gaussian Channel $EZ^2 = N$, $EZ = 0$, so

$$\begin{aligned} EY^2 &= E(X+Z)^2 \\ &= EX^2 + 2EXEZ + EZ^2 \\ &= P + N \end{aligned}$$

Since when $EY^2 = P + N$, $h(Y)$ is bounded by $\frac{1}{2}\log 2\pi e(P+N)$, so

$I(X;Y) = h(Y) - h(Z)$, equal hold when $X \sim N(0, P)$

$$\begin{aligned} &\le \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN \\ &= \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \end{aligned}$$

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

Compression: From the above result, we know the capacity of a power-limited Gaussian channel. In compression, we trying to use minimum bits to transmit the information and achieve a high rate. When information is transmitted over a power limited channel, the rate can' be above the capacity.

System identification: We want to estimate the pdf of an unknown distribution. We may get a solution by maximizing the entropy h(f) over all pdf f satisfying some moment constraints. $f(x) = e^{\lambda_0 - 1 + \sum_{i=1}^{m} \lambda_i r_i(x)}$ where $\lambda_0, \lambda_1 \ldots \lambda_m$ are chosen so that f satisfies the constraints

Spectrum estimation: The entropy rate of a stochastic process subject to autocorrelation constraints $r_0, r_1, \ldots r_p$ is maximized by the pth order zero- mean Gauss-Markov process satisfying these constraints. The maximum entropy spectrum is:

$$S(l) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} \alpha_k e^{-ikl}\right|^2}$$

(d) Explain Burg's Maximum Entropy Theorem

Burg's Maximum Entropy Theorem:

the maximum entropy rate stochastic process, satisfying the constraints, $EX_iX_{i+k} = a_k, k = 0, 1, 2 \ldots p$ for all i, is the pth order Gauss- Markov process of the form $X_i = -\sum_{k=1}^{p} a_k X_{i-k} + Z_i$ where the Z_i are $iid \sim N(0, \sigma^2)$, and , $\alpha_1, \alpha_2 \ldots \alpha_p, \sigma^2$ are chosen to satisfy the constraints $EX_iX_{i+k} = a_k$

The basic assumption involved in maximum entropy spectrum analysis is that the stationary time series being analyzed is the most random or the least predicable ones. The maximum entropy time series is governed by a Gaussian joint probability function.

The entropy rate of a stochastic process subject to auto correlation constraints $r_0, r_1, \ldots r_p$ is maximized by the pth order zero- mean Gauss-Markov process satisfying these constraints. The maximum entropy spectrum is:

$$S(l) = \frac{\sigma^2}{\left|1 + \sum_{k=1}^{p} \alpha_k e^{-ikl}\right|^2}$$

In a practical problem, we calculate the autocorrelation of a given sample sequences, choosing a optimal value p and get the estimated spectrum we wanted.

Problem No. 3: Statistics

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

(a) Describe the type class for this event.

P(1) = 1/9;
P(2) = 2/9;
P(3) = 1/9;
P(4) = 2/9;
P(5) = 1/9;
P(6) = 2/9;

This type class can be described as the sequences of length 9 and including 1, 2,3,4, 5,6 with the above probability.

(b) Bound the size of the type class.

$$\frac{1}{(n+1)^{|x|}} 2^{nH(p)} \le |T(p)| \le 2^{nH(p)}$$

here n=9, $|x| = 6$, H(p)= H(1/9, 2/9, 1/9, 2/9, 1/9, 2/9).

(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.

1) We can use $Q^n(T(P)) \approx 2^{-n(D(P||Q) + H(p))}$ to estimate the probability. Here Q is the true distribution and P is the distribution got from samples. If this is not a fair die, to get the true distribution of numbers, we need to have large amount of data. We can find even if the distance between P and Q are not big, since it's on the exponential part and timed by n, the probability will be very small.

2) We can also use $2^{-nH(P)}$, this correspond to the case when x is in the type class of Q. Now the probability is decided by the length of the sequences and the entropy of the true distribution

3) For some case, the die may be a fair one, we can also calculate the probability of this sequence by multiplying the probabilities of each independent events.