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

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \sum p(x) H(Y|(X = x))$$

$$= H(Y) - \sum p(x) H(p)$$

$$= H(Y) - H(p)$$

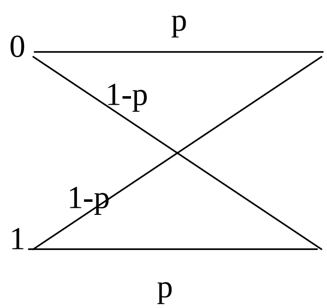
$$I(X;Y) \leq 1 - H(p)$$

(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.

First, because we already know the prior distribution of $p(x)$, we can construct a huffman code to minimize the number of total bits transmitted. Then, according to the channel coding theory, it must exist a sequence of $(2^{nR}, n)$ codes with maximum probability of error converges to zero, if $R < C$, such as the Hamming codes. In general, from the joint source channel coding theorem we know we can consider the design of a communication system as a combination of two parts, source coding and channel coding The combination will be as efficient as anything we could design by considering both problems together.

(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.

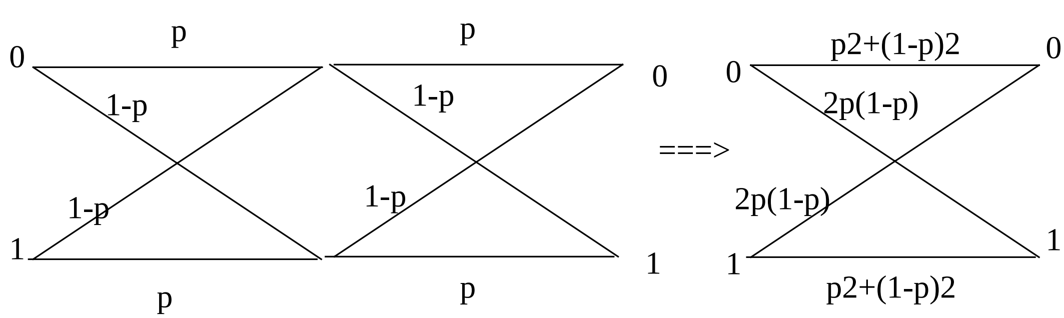
For a single BSC,



the capacity can be computed by:

$$C = 1 - H(p)$$

In order to find the expression of the capacity of the cascade of two BSC, we combine two channels into one:



The capacity can be computed by:

$$C = 1 - H(p^2 + (1-p)^2)$$

The capacity decreases compared to one single BSC. Because the more channels we cascade, the more errors we have, and the mutual information of x and y will decrease, so does the overall capacity.

Problem No. 2: Continuous Random Variables

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

Let's assume $Y = AX$, so $f(x) = \frac{1}{|A|} f(A^{-1}y)$

$$H(Ax) = -\int \frac{1}{|A|} f(A^{-1}y) \ln \frac{1}{|A|} f(A^{-1}y) dy$$

$$= -\int f(x) \ln \frac{1}{|A|} dx - \int f(x) \ln f(x) dx \quad \text{(use } f(x) = \frac{1}{|A|} f(A^{-1}y)\text{)}$$

$$= \ln(|A|) + H(x) \qquad \text{(use definition and } \int f(x) dx = 1\text{)}$$

(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).

The expression is $C = max I(X;Y) \, (p(x) : EX^2 < P)$

Expanding $I(X;Y)$, we have:

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

we know $H(Z) = \frac{1}{2}\log 2\pi eN$. Also:

$$EY^2 = E(X+Z)^2 = EX^2 + 2EXEZ + EZ^2 = P + N$$

and Y is bounded by $\frac{1}{2}\log 2\pi e(P+N)$, so we can obtain

$$I(X;Y) = H(Y) - H(Z)$$
$$<= \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN$$
$$= \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

and the maximum is attained when $X \sim N(0, P)$.

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

Compression: If we apply power limit on channel, we cannot achieve the highest compression rate R.

System identification: Objective is to distribute the total power among the channels so as to maximize the capacity.

Spectrum : The entropy rate of a stochastic process subject to autocorrelation constraints $R_1, R_2 ..., R_p$ is maximized by the pth order zero-mean Gauss-Markov process satisfying these constraints. The maximum entropy spectrum is:

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

(d) Explain Burg's Maximum Entropy Theorem

In some applications, we can assume an underlying autoregressive model for the data, this method (called Burg's method) has proved useful in determining the parameters of the model. It's a powerful method for estimation of spectral densities.

The theorem is: the maximum entropy rate stochastic process $\{X_i\}$,satisfying the constraints $EX_iX_{i+k} = a_k,\ k = 0, 1, 2 \ldots, p$, for all i, is the pth order Gauss-Markow process of the form $X_i = -\sum_{k=1}^{p} a_k X_{i-k} + Z_i$ where the Z_i are $i, i, d \sim N(0, \alpha^2)$ and $a_1, a_2, \ldots a_p,\ a^2$ are chosen to satisfy the constraints given previously.

Note that in Burg's theory, we don't assume that $\{X_i\}$ is zero mean or Gaussian or wide-sense stationary.

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.

The length of sequence is 9, P(a) are given by:

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

The type class of this event is the set of all sequences of length 9 with one 1, two 2's, one 3, two 4's, one 5 and two 6's

(b) Bound the size of the type class.

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

Where $n = 9$ and $|x| = 6$ in our case.

(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) Use $Q^n(T(P)) \approx 2^{-nD(P||Q)}$ to compute the probability of this event $P(x) = \left\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right\}$. and $Q(x) = \left\{\frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}\right\}$. The relative entropy $D(P||Q)$ is computed as $\sum_x p(x)\log\frac{p(x)}{q(x)} = \log\frac{3}{2\sqrt{2}}$. So the probability of this event $Q^n(T(P))$ is computed as $2^{-nD(P||Q)}$. which equals to $\left(\frac{2\sqrt{2}}{3}\right)^9$

2) We can use $2^{-nH(P)}$ to approximate to probability of the event, where the $H(P)$ is equal to $Q(x)$ in method (1). So $H(P) = H\left(\frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}\right) = \log\left(\frac{9}{\sqrt[3]{4}}\right)$, then we can compute the $2^{-nH(P)}$, which is equal to $\left(\frac{\sqrt[3]{4}}{9}\right)^9$.

3) We can treat the event as a sequence of independent event, thus $P(X^n) = P(X_1)P(X_2)...P(X_9)$, which is equal to $\left(\frac{2}{9}\right)^6\left(\frac{1}{9}\right)^3$.