Name: Suresh Balakrishnama

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

Solution:

In a binary symmetric channel when an error occurs, a 0 is received as a 1 and vice versa. The received bits do not reveal where the errors have occurred. A simple binary symmetric channel is shown in the figure 1 in which the input symbols are complemented with probability p. This is the simplest model of a channel with errors but it captures the complexity involved in a general channel.

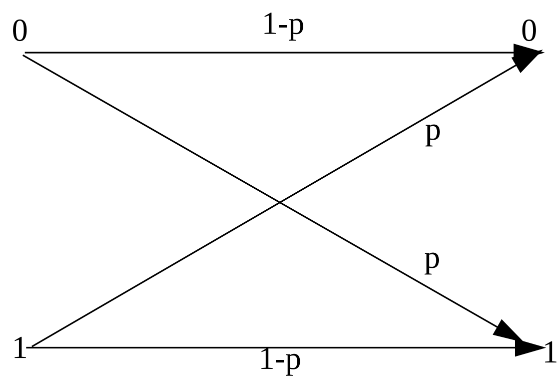


Figure1: A Binary symmetric channel

Bound on the mutual information is given by

$$I(X;Y) = h(Y) - h\langle Y|X\rangle$$

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

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

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

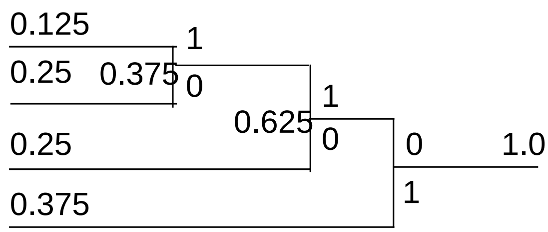
$$I(X;Y) \le 1 - H(p)$$

(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.

Solution:

Since we know the prior probability information about the letters, we can use Huffman code to find out the average code length required for transmitting.

Construction of Huffman code,



Using Huffman code,

Symbol	**P(A)**	**codeword**	**length**
a	0.25	010	3
b	0.25	00	2
c	0.125	011	3
d	0.375	1	1

Average Codeword Length = $0.25 \times 3 + 0.25 \times 2 + 0.125 \times 3 + 0.375 \times 1$

$$= 2bits$$

Therefore, to transmit the given symbols we require at the minimum $2bits$ to transmit with minimum errors.

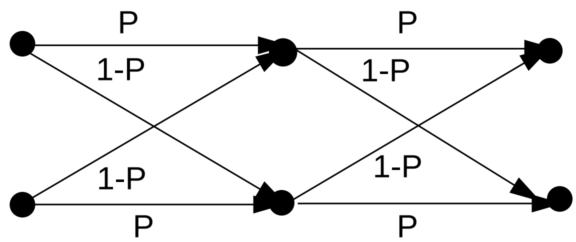
(c) Suppose we add a second BSC with the same properties as the first. Derive an expression for the capacity, state whether the capacity increases or decreases, and explain why.

Solution:

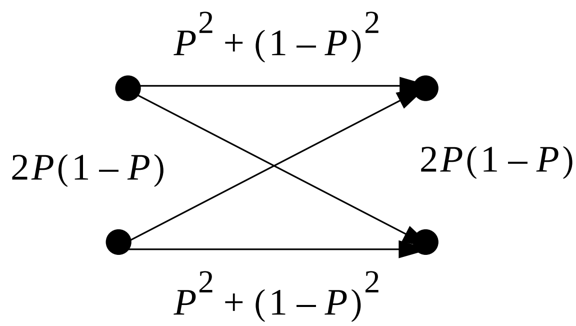
The capacity of the cascade of two channels is

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

Two Binary Symmetric channels connected in cascade is shown below:



The equivalent as a single symmetric channel is



The capacity decreases because when we cascade more and more BSC, we are correspondingly increasing the error associated with each channel. Since we know Binary symmetric channel is the simplest model of a channel with errors.

Problem No. 2:

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

Solution:

Scaling theorem provides the proof that translation does not change the differential entropy.

$$h(AX) = \log|det(A)| + h(X)$$

Let $Y = AX$, then $f_Y(y) = \frac{1}{|A|}f_x\left(\frac{y}{A}\right)$ and

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

Hence proved.

(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).

Solution:

The information capacity of the Gaussian channel with power constraint P is

$$C = max_{EX^2 \le P} I(X;Y)$$

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

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

since Z is independent of X. Now, $h(Z) = \frac{1}{2}\log 2\pi eN$ and also,

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

since X and Z are independent and $EZ = 0$. Given $EY^2 = P + N$, the entropy of Y is bounded by $\frac{1}{2}\log 2\pi e(P+N)$

Applying this result to bound the mutual information, we obtain

$$I(X;Y) = h(Y) - h\langle Y|X\rangle$$

$$I(X;Y) \le \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi eN$$

$$= \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

Hence, the information capacity of the Gaussian channel is

$$C = max_{EX^2 \le P} I(X;Y) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right)$$

and the maximum is obtained 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.

Solution:

As an example of system identification let us consider a communication system with band-limited channel. For infinite bandwidth channels, the capacity grows linearly with power. Accordingly, the range of rate also increases. With power constraint the channel capacity can also increase.

To obtain a better compression scheme, we know that $(2^{nR},n)$ code exist with $R < C$, we can further distribute this into good and worst codeword using the power constraint. The codewords not satisfying the power constraint have probability of error 1 and belong to the worst half of the codewords. This worst set of codewords can be deleted from the set of code for transmission.

(d) Explain Burg's Maximum Entropy Theorem.

Solution:

The maximum entropy rate stochastic process X_i satisfying the constraints $EX_iX_{i+k} = \alpha_k$ for $k = 0,1\ldots\ldots,p$ 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 Z_i are i.i.d and a_k, σ^2 are chosen to satisfy the constraint equation.

According to Burg's maximum entropy theorem, entropy of a segment of a stochastic process is bounded by entropy of segment of a Gaussian process with the same covariance values. The covariance values are the constraint between the stochastic process and Gaussian process. Given the process we can choose the values of a_k to satisfy the constraints and in this process we use the *Yule-Walker* equations.

Application: Burg's theorem finds importance in spectral estimation. The simplest way to estimate the power spectrum is to estimate the autocorrelation function by taking the sample average for sample of length n. However, the estimate obtained from this periodogram method does not converge to the true power spectrum especially if n is large. The main problem with this periodgram method is that the estimates of the autocorrelation function from the data have different accuracies. The estimates for low values of lags are based on a large number of samples and vice versa. The estimates are more accurate on low values of lags since size of samples is large. Since higher values of lags are not so useful the autocorrelation functions at higher lags can be set to zero. But this introduces inaccuracies because of sudden shift to zero value. Burg introduced a more consistent alternative method. All the autocorrelation functions at high lags were set to the values that maximize the entropy rate of the process. This is the reason his theorem is associated with maximum entropy.

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.

Solution:

The length of sequence given in this case is 9 i.e., 9 events occur on the roll of a die. There are some occurrences which are repeated.

Let us calculate the probability of each occurrence. Let X be the variable representing the outcome of each roll of a die.

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

The type class or empirical probability distribution of a sequence $x_1, x_2 \ldots\ldots, x_n$ is the relative proportion of occurrences of each symbol from the alphabet.

Hence, for our problem the type class of the sequence of length 9 is occurrence of 1, 3, 5 once and 2, 4, 6 twice.

(b) Bound the size of the type class.

Solution:

Size of a type class $T(P)$ is given by:

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

In the problem given $n = 9$

(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.

Solution:

(a) Probability of type class:

$$Q^n(T(P)) = D^{-nD\langle P||Q\rangle}$$

Here, $Q = \{\frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}\}$ is the probability for each of the events in the given sequence and,

$P = \frac{1}{6}$ for each of the outcome on the roll of die. Since rolling of a die is a uniform distribution the probability for any face of the die is same.

Using the values of Q and P, the probability of the event can be calculated.

(b) Approximation using Sanov's theorem

We can estimate the probability closest to sequence Q by,

$$P^* = \frac{Q(x)e^{\left(\sum_i \lambda_i\right) g_i(x)}}{\sum_{a \in \chi} Q(a)e^{\left(\sum_i \lambda_i\right) g_i(x)}}$$