Applied Information Theory 650

`http://www.cs.umass.edu/~elm/Teaching/650_F14/`

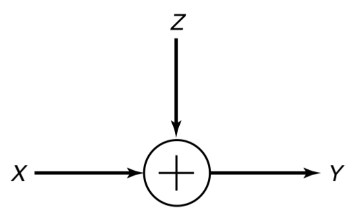
Assignment 5

7.1 *Preprocessing the output.* (5 points) One is given a communication channel with transition probabilities $p(y|x)$ and channel capacity $C = max_{p(x)} I(X;Y)$. A helpful statistician preprocesses the output by forming $\widetilde{Y} = g(Y)$. He claims that this will strictly improve the capacity.

(a) Show that he is wrong.

(b) Under what conditions does he not strictly decrease the capacity?

7.2 *Additive noise channel.* (10 points) Find the channel capacity of the following discrete memoryless channel:



where $Pr\{Z = 0\} = Pr\{Z = a\} = \frac{1}{2}$. The alphabet for x is $\mathcal{X} = \{0, 1\}$. Assume that Z is independent of X. Observe that the channel capacity depends on the value of a.

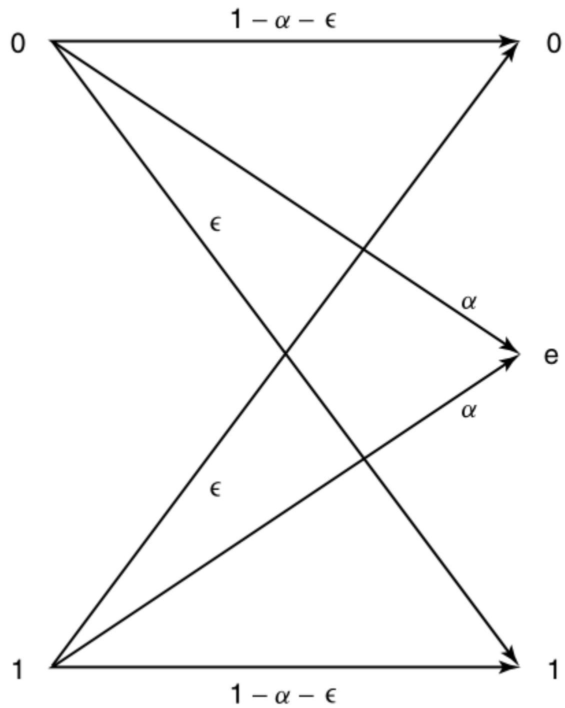
7.3 *Channels with memory have higher capacity.* (10 points) Consider a binary symmetric channel with $Y_i = X_i \oplus Z_i$, where $\oplus$ is mod 2 addition, and $X_i, Y_i \in \{0, 1\}$. Suppose that $\{Z_i\}$ has constant marginal probabilities $Pr\{Z_i = 1\} = p = 1 - Pr\{Z_i = 0\}$, but that Z_1 , Z_2 , ..., Z_n are not necessarily independent. Assume that Z_n is independent of the input X_n . Let $C = 1 - H(p, 1 - p)$. Show that $max_{p(x_1,x_2,...,x_n)} I(X_1, X_2, ..., X_n; Y_1, Y_2, ..., Y_n) \geq nC$.

7.8 *Z-channel.* (10 points) The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad x, y \in \{0, 1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

7.13 *Erasures and errors in a binary channel.* NOTE: THis is a tough problem. I've made it worth 5 points so it won't affect your grade too much if you can't get it. (5 points) Consider a channel with binary inputs that has both erasures and errors. Let the probability of error be ϵ and the probability of erasure be α, so the channel is as follows:

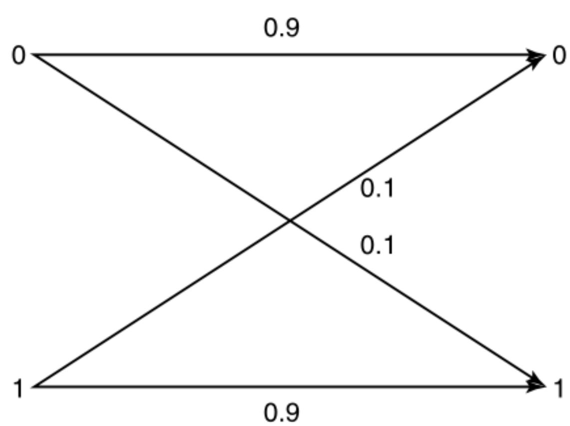


(a) Find the capacity of this channel.

(b) Specialize to the case of the binary symmetric channel ($\alpha = 0$).

(c) Specialize to the case of the binary erasure channel ($\epsilon = 0$).

7.15 *Jointly typical sequences.* (5 points for each part) As we did in Problem 3.13 for the typical set for a single random variable, we will calculate the jointly typical set for a pair of random variables connected by a binary symmetric channel, and the probability of error for jointly typical decoding for such a channel.



We consider a binary symmetric channel with crossover probability 0.1. The input distribution that achieves capacity is the uniform distribution [i.e., $p(x) = (\frac{1}{2}, \frac{1}{2})$], which yields the joint distribution $p(x, y)$ for this channel is given by

		Y	
		0	1
X	0	0.45	0.05
	1	0.05	0.45

The marginal distribution of Y is also $(\frac{1}{2}, \frac{1}{2})$.

(a) Calculate $H(X)$, $H(Y)$, $H(X, Y)$, and $I(X; Y)$ for the joint distribution above.

(b) Let $X_1, X_2, ..., X_n$ be drawn i.i.d. according the Bernoulli($\frac{1}{2}$) distribution. Of the 2^npossible input sequences of length n, which of them are typical [i.e., member of $A_\epsilon^{(n)}(X)$ for $\epsilon = 0.2$]? Which are the typical sequences in $A_\epsilon^{(n)}(Y)$?

(c) The jointly typical set $A_\epsilon^{(n)}(X, Y)$ is defined as the set of sequences that satisfy equations (7.35-7.37). The first two equations correspond to the conditions that x^n and y^n are in $A_\epsilon^{(n)}(X)$ and $A_\epsilon^{(n)}(Y)$, respectively. Consider the last condition, which can be rewritten to state that $-\frac{1}{n}$ log $p(x^n, y^n) \in (H(X, Y) - \epsilon, H(X, Y) + \epsilon)$. Let k be the number of places in which the sequence x^n differs from y^n (k is a function of the two sequences). Then we can write

$$p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i) \tag{7.156}$$

$$= (0.45)^{n-k}(0.05)^k \tag{7.157}$$

$$= \left(\frac{1}{2}\right)^n (1-p)^{n-k}p^k. \tag{7.158}$$

An alternative way at looking at this probability is to look at the binary symmetric channel as an additive channel $Y = X \oplus Z$, where Z is a binary random variable that is equal to 1 with probability p, and is independent of X. In this case,

$$p(x^n, y^n) = p(x^n)p(y^n|x^n) \tag{7.159}$$

$$= p(x^n)p(z^n|x^n) \tag{7.160}$$

$$= p(x^n)p(z^n) \tag{7.161}$$

$$= \left(\frac{1}{2}\right)^n (1-p)^{n-k}p^k. \tag{7.162}$$

Show that the condition that (x^n, y^n) being jointly typical is equivalent to the condition that x^n is typical and $z^n = y^n - x^n$ is typical.

(d) We now calculate the size of $A_\epsilon^{(n)}(Z)$ for $n = 25$ and $\epsilon = 0.2$. As in Problem 3.13, here is a table of the probabilities and numbers of sequences with k ones:

k	$\binom{n}{k}$	$\binom{n}{k}p^k(1-p)^{n-k}$	$-\frac{1}{n}\log p(x^n)$
0	1	0.071790	0.152003
1	25	0.199416	0.278800
2	300	0.265888	0.405597
3	2300	0.226497	0.532394
4	12650	0.138415	0.659191
5	53130	0.064594	0.785988
6	177100	0.023924	0.912785
7	480700	0.007215	1.039582
8	1081575	0.001804	1.166379
9	2042975	0.000379	1.293176
10	3268760	0.000067	1.419973
11	4457400	0.000010	1.546770
12	5200300	0.000001	1.673567

[Sequences with more than 12 ones are omitted since their total probability is negligible (and they are not in the typical set).]

What is the size of the set $A_\epsilon^{(n)}(Z)$?

(e) Now consider random coding for the channel, as in the proof of the channel coding theorem. Assume that 2^{nR} codewords $X^n(1), X^n(2), ..., X^n(2^{nR})$ are chosen uniformly over the 2^n possible binary sequences of length n. One of these codewords is chosen and sent over the channel. The receiver looks at the received sequence and tries to find a codeword in the code that is jointly typical with the received sequence. As argued above, this corresponds to finding a codeword $X^n(i)$ such that $Y^n - X^n(i) \in A_\epsilon^{(n)}(Z)$. For a fixed codeword $x^n(i)$, what is the probability that the received sequence Y^n is such that $(x^n(i), Y^n)$ is jointly typical?

(f) Now consider a particular received sequence $y^n = 000000...0$, say. Assume that we choose a sequence X^n at random, uniformly distributed among all the 2^n possible binary n-sequences. What is the probability that the chosen sequence is jointly typical with this y^n? [Hint: This is the probability of all sequences x^n such that $y^n - x^n \in A_\epsilon^{(n)}(Z)$.]

(g) Now consider a code with $2^9 = 512$ codewords of length 12 chosen at random, uniformly distributed among all the 2^n sequences of length $n = 25$. One of these codewords, say the one corresponding to $i = 1$, is chosen and sent over the channel. As calculated in part (e), the received sequence, with high probability, is jointly typical with the codeword that was sent. What is the probability that one or more of the other codewords (which were chosen at random, independent of the sent codeword) is jointly typical with the received sequence?

[*Hint*: You could use the union bound, but you could also calculate this probability exactly, using the result of part (f) and the independence of the codewords.]

(h) Given that a particular codeword was sent, the probability of error (averaged over the probability distribution of the channel and over the random choice of other codewords) can be written as

$$\Pr(\text{Error})|x^n(1) \text{ sent}) = \sum_{y^n : y^n \text{causes error}} p(y^n|x^n(1)).$$

There are two kinds of error: the first occurs if the received sequence y^n is not jointly typical with the transmitted codeword, and the second occurs if there is another codeword jointly typical with the received sequence. Using the result of the preceding parts, calculate this probability of error. By the symmetry of the random coding argument, this does not depend on which codeword was sent.

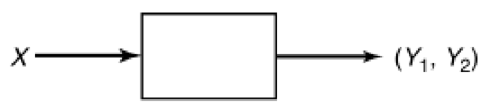
The calculations above show that average probability of error for a random code with 512 codewords of length 25 over the binary symmetric channel of crossover probability 0.1 is about 0.34. This seems quite high, but the reason for this is that the value of ϵ that we have chosen is too large. By choosing a smaller ϵ and a larger n in the definitions of $A_\epsilon^{(n)}$, we can get the probability of error to be as small as we want as long as the rate of the code is less than $I(X;Y) - 3\epsilon$.

Also note that the decoding procedure described in the problem is not optimal. The optimal decoding procedure is maximum likelihood (i.e., to choose the codeword that is closest to the received sequence). It is possible to calculate the average probability of error for a random code for which the decoding is based on an approximation to maximum likelihood decoding, where we decode a received sequence to the unique codeword that differs from the received sequence in 4 bits, and declare an error otherwise. The only difference with the jointly typical decoding described above is that in the case when the codeword is equal to the received sequence! The average probability of error for this decoding scheme can be shown to be about 0.285.

7.20 *Channel with two independent looks at Y.* (10 points) Let Y_1 and Y_2 be conditionally independent and conditionally identically distributed given X.

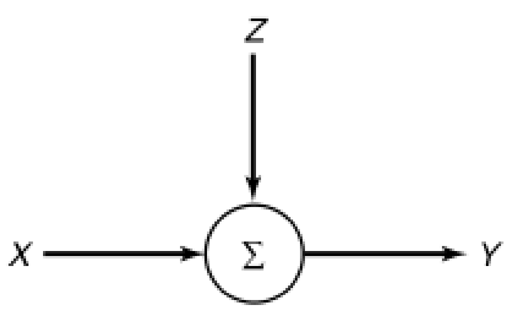
(a) Show that $I(X;Y_1,Y_2) = 2I(X;Y_1) - I(Y_1;Y_2)$.

(b) Conclude that the capacity of the channel



is less than twice the capacity of the channel

7.30 *Noise alphabets.* (10 points) Consider the channel



$\mathcal{X} = \{0,1,2,3\}$, where $Y = X + Z$, and Z is uniformly distributed over three distinct integer values $\mathcal{Z} = \{z_1, z_2, z_3\}$.

(a) What is the maximum capacity over all choices of the $\mathcal{Z}$ alphabet? Give distinct integer values z_1, z_2, z_3 and a distribution on $\mathcal{X}$ achieving this.

(b) What is the minimum capacity over all choices for the $\mathcal{Z}$ alphabet? Give distinct integer values z_1, z_2, z_3 and a distribution on $\mathcal{X}$ achieving this.