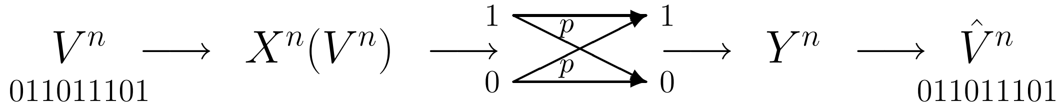
EE 376A
Information Theory
Prof. T. Cover

Handout #30
Thursday, March 12, 2009

Solutions to Homework Set #8

1. **Source and channel.**
   We wish to encode a Bernoulli(α) process $V_1, V_2, \ldots$ for transmission over a binary symmetric channel with error probability p.



   Find conditions on α and p so that the probability of error $P(\hat{V}^n \neq V^n)$ can be made to go to zero as $n \longrightarrow \infty$.

   Solution: Source and channel.
   Suppose we want to send a binary i.i.d. Bernoulli(α) source over a binary symmetric channel with error probability p.

   By the source-channel separation theorem, in order to achieve the probability of error that vanishes asymptotically, i.e. $P(\hat{V}^n \neq V^n) \to 0$, we need the entropy of the source to be less than the capacity of the channel. Hence,

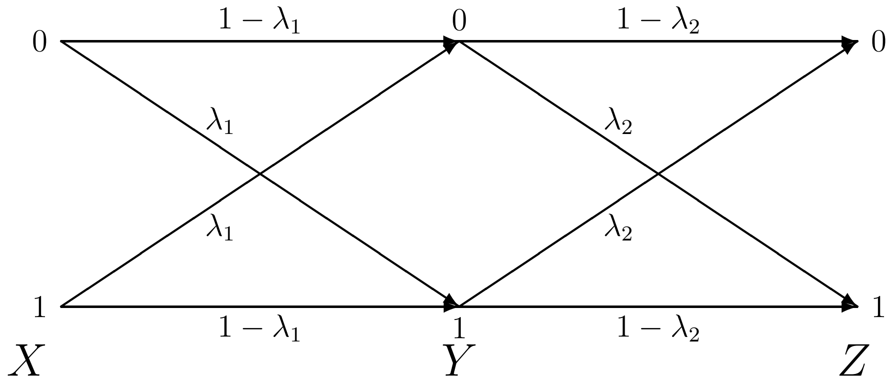
$$H(\alpha) + H(p) < 1,$$

   or, equivalently,

$$\alpha^\alpha (1-\alpha)^{1-\alpha} p^p (1-p)^{1-p} > \frac{1}{2}.$$

2. **Cascaded BSCs.**
   Consider the two discrete memoryless channels $(\mathcal{X}, p_1(y|x), \mathcal{Y})$ and $(\mathcal{Y}, p_2(z|y), \mathcal{Z})$.

   Let $p_1(y|x)$ and $p_2(z|y)$ be binary symmetric channels with crossover probabilities λ_1 and λ_2 respectively.



(a) What is the capacity C_1 of $p_1(y|x)$?

(b) What is the capacity C_2 of $p_2(z|y)$?

(c) We now cascade these channels. Thus $p_3(z|x) = \sum_y p_1(y|x)p_2(z|y)$. What is the capacity C_3 of $p_3(z|x)$? Show $C_3 \le \min\{C_1, C_2\}$.

(d) Now let us actively intervene between channels 1 and 2, rather than passively transmitting y^n. What is the capacity of channel 1 followed by channel 2 if you are allowed to decode the output y^n of channel 1 and then reencode it as $\tilde{y}^n$ for transmission over channel 2? (Think $W \longrightarrow x^n(W) \longrightarrow Y^n \longrightarrow \tilde{y}^n(Y^n) \longrightarrow Z^n \longrightarrow \hat{W}$.)

(e) What is the capacity of the cascade in part c) if the receiver can view *both* Y and Z?

Solution: Cascaded BSCs

(a) C_1 is just a capacity of a $BSC(\lambda_1)$. Thus, $C_1 = 1 - H(\lambda_1)$.

(b) Similarly, $C_2 = 1 - H(\lambda_2)$.

(c) $X \to Y \to Z$ forms a Markov chain. From data processing inequality,

$$I(X;Z) \le I(X;Y)$$

and

$$I(X;Z) \le I(Y;Z).$$

Hence, $C \le \min\{C_1, C_2\}$.

One can show that $C \le \min\{C_1, C_2\}$ in a more direct way. First observe that the cascaded channel is also a BSC. Since the new BSC has a crossover probability of $p_3 = \lambda_1(1-\lambda_2) + (1-\lambda_1)\lambda_2 = \lambda_1 + \lambda_2 - 2\lambda_1\lambda_2$,

$$C_3 = 1 - H(\lambda_1 + \lambda_2 - 2\lambda_1\lambda_2).$$

Note that the new channel is noisier than the original two since by concavity of $H(p)$,

$$H((1-\lambda_1)\lambda_2 + \lambda_1(1-\lambda_2)) \geq \lambda_2 H(1-\lambda_1) + (1-\lambda_2)H(\lambda_1) = H(\lambda_1)$$

Similarly for $H(\lambda_2)$. Thus, $C_3 \leq \min\{C_1, C_2\}$.

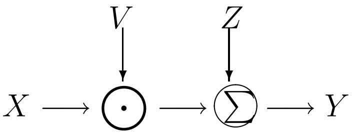
(d) Since we are allowed to decode the intermediate outputs and reencode them prior to the second transmission, any rate less than both C_1 and C_2 can be achievable and at the same time any rate greater than either C_1 or C_2 will cause $P_e^{(n)} \to 1$ exponentially. Hence, the overall capacity is the minimum of two capacities,

$$\min(C_1, C_2) = \min(1 - H(\lambda_1), 1 - H(\lambda_2)).$$

(e) Note that Z becomes irrelevant once we observe Y. Thus, the capacity of this channel is just $C_1 = 1 - H(\lambda_1)$.

Alternatively, $X \to Y \to (Y, Z)$ forms a Markov chain so that $I(X;Y) \geq I(X;Y,Z)$. On the other hand, $I(X;Y) \leq I(X;Y,Z)$ since we can always ignore the observation Z. (Or $X \to (Y,Z) \to Y$ also forms a Markov chain.) Hence, $I(X;Y) = I(X;Y,Z)$ and the capacity of this case is C_1.

3. **Fading channel.**
Consider an additive noise fading channel



$$Y = XV + Z,$$

where Z is additive noise, V is a random variable representing fading, and Z and V are independent of each other and of X.

(a) Argue that knowledge of the fading factor V improves capacity by showing

$$I(X;Y|V) \geq I(X;Y).$$

(b) Incidentally, conditioning does not always increase mutual information. Give an example of $p(u,r,s)$ such that $I(U;R|S) < I(U;R)$.

Solution: Fading channel

(a) We may show the inequality as follows:

$$\begin{aligned} I(X;Y|V) &= h(X|V) - h(X|Y,V) \\ &= h(X) - h(X|Y,V) \qquad (1) \\ &\geq h(X) - h(X|Y) \qquad (2) \\ &= I(X;Y) \end{aligned}$$

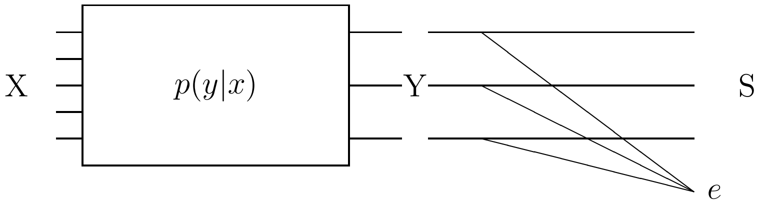
where (1) follows from the independence of X and V, and (2) is true because conditioning reduces entropy.

(b) The inequality can go the other direction as well. If U is drawn according to $\mathrm{Ber}(\frac{1}{2})$ and $U = R = S$, we have $I(U;R) = 1$, but $I(U;R|S) = 0$, thus giving

$$I(U;R|S) < I(U;R).$$

4. **Erasure channel**

Let $\{\mathcal{X}, p(y|x), \mathcal{Y}\}$ be a discrete memoryless channel with capacity C. Suppose this channel is immediately cascaded with an erasure channel $\{\mathcal{Y}, p(s|y), \mathcal{S}\}$ that erases α of its symbols.



Specifically, $\mathcal{S} = \{y_1, y_2, \ldots, y_m, e\}$, and

$$\begin{aligned} \Pr\{S = y|X = x\} &= \bar{\alpha}p(y|x), \quad y \in \mathcal{Y}, \\ \Pr\{S = e|X = x\} &= \alpha. \end{aligned}$$

Determine the capacity of this channel.

Solution: Erasure channel

The capacity of the channel is

$$C = \max_{p(x)} I(X;S) \qquad (3)$$

Define a new random variable Z, a function of S, where $Z = 1$ if $S = e$ and $Z = 0$ otherwise. Note that $p(Z = 1) = \alpha$ independent of X. Expanding the mutual information,

$$\begin{aligned} I(X;S) &= H(S) - H(S|X) && (4)\\ &= H(S,Z) - H(S,Z|X) && (5)\\ &= H(Z) + H(S|Z) - H(Z|X) - H(S|X,Z) && (6)\\ &= I(X;Z) + I(S;X|Z) && (7)\\ &= 0 + \alpha I(X;S|Z=1) + (1-\alpha) I(X;S|Z=0) && (8) \end{aligned}$$

When $Z = 1$, $S = e$ and $H(S|Z = 1) = H(S|X, Z = 1) = 0$. When $Z = 0$, $S = Y$, and $I(X; S|Z = 0) = I(X; Y)$. Thus

$$I(X;S) = (1-\alpha) I(X;Y) \tag{9}$$

and therefore the capacity of the cascade of a channel with an erasure channel is $(1-\alpha)$ times the capacity of the original channel.

5. **Random "20" questions**
Let X be uniformly distributed over $\{1, 2, \ldots, m\}$. Assume $m = 2^n$. We ask random questions: Is $X \in S_1$? Is $X \in S_2$?...until only one integer remains. All 2^m subsets S of $\{1, 2, \ldots, m\}$ are equally likely. The questions are independently and identically distributed, and subsets are drawn with replacement.

   (a) Firstly, how many deterministic questions would be needed to determine X?

   (b) Henceforth, we generate questions randomly in the manner described above. Without loss of generality, suppose that $X = 1$ is the random object. What is the probability that object 2 yields the same answers for k random questions as object 1?

   (c) What is the expected number of objects in $\{2, 3, \ldots, m\}$ that have the same answers to k random questions as does the correct object 1?

   (d) Suppose we ask $n + \sqrt{n}$ random questions. What is the expected number of wrong objects agreeing with the answers?

   (e) Use Markov's inequality $\Pr\{N \geq t\} \leq \frac{EN}{t}$, to show that the probability of error (one or more wrong object remaining) goes to zero as $n \longrightarrow \infty$.

Solution: Random "20" questions.

(a) Obviously, Huffman codewords for X are all of length n. Hence, with n deterministic questions, we can identify an object out of 2^n candidates.

(b) Observe that the total number of subsets which include both object 1 and object 2 or neither of them is 2^{m-1}. Hence, the probability that object 2 yields the same answers for k questions as object 1 is $(2^{m-1}/2^m)^k = 2^{-k}$.

More information theoretically, we can view this problem as a channel coding problem through a noiseless channel. Since all subsets are equally likely, the probability the object 1 is in a specific random subset is $1/2$. Hence, the question whether object 1 belongs to the kth subset or not corresponds to the kth bit of the random codeword for object 1, where codewords X^k are Bern(1/2) random k-sequences.

Object	Codeword
1	$0110\dots1$
2	$0010\dots0$
$\vdots$	

Now we observe a noiseless output Y^k of X^k and figure out which object was sent. From the same line of reasoning as in the achievability proof of the channel coding theorem, i.e. joint typicality, it is obvious the probability that object 2 has the same codeword as object 1 is 2^{-k}.

(c) Let

$$1_j = \begin{cases} 1, & \text{object } j \text{ yields the same answers for } k \text{ questions as object 1} \\ 0, & \text{otherwise} \end{cases},$$

for $j = 2, \dots, m$.

Then,

$$\begin{aligned} E(\#\text{ of objects in } \{2,3,\dots,m\} \text{ with the same answers}) &= E(\sum_{j=2}^{m} 1_j) \\ &= \sum_{j=2}^{m} E(1_j) \\ &= \sum_{j=2}^{m} 2^{-k} \\ &= (m-1)2^{-k} \\ &= (2^n-1)2^{-k}. \end{aligned}$$

(d) Plugging $k = n + \sqrt{n}$ into (c) we have the expected number of $(2^n-1)2^{-n-\sqrt{n}}$.

(e) Let N by the number of wrong objects remaining. Then, by Markov's inequality

$$P(N \geq 1) \;\leq\; EN = (2^n-1)2^{-n-\sqrt{n}} \leq 2^{-\sqrt{n}} \to 0,$$

where the first equality follows from part (d).