EE276 Homework #5 Solutions

Due on Friday Feb 20, 6pm - Gradescope entry code: E6VP4X

1. **Zero-error capacity.** A channel with alphabet $\{0, 1, 2, 3, 4\}$ has transition probabilities of the form

$$p(y|x) = \begin{cases} 1/2 & \text{if } y = x \pm 1 \bmod 5 \\ 0 & \text{otherwise.} \end{cases}$$

   (a) Compute the capacity of this channel in bits.

   (b) The zero-error capacity of a channel is the number of bits per channel use that can be transmitted with zero probability of error. Clearly, the zero-error capacity of this pentagonal channel is at least 1 bit (transmit 0 or 1 with probability 1/2). Find a block code that shows that the zero-error capacity is greater than 1 bit. Can you estimate the exact value of the zero-error capacity?

   (*Hint*: Consider codes of length 2 for this channel.)

Solution: *Zero-error capacity.*

(a) Since the channel is symmetric, it is easy to compute its capacity:

$$\begin{aligned} H(Y|X) &= 1 \\ I(X;Y) &= H(Y) - H(Y|X) = H(Y) - 1\,. \end{aligned}$$

So mutual information is maximized when Y is uniformly distributed, which occurs when the input X is uniformly distributed. Therefore the capacity in bits is $C = \log_2 5 - 1 = \log_2 2.5 = 1.32$.

(b) Let us construct a block code consisting of 2-tuples. We need more than 4 codewords in order to achieve capacity greater than 2 bits, so we will pick 5 codewords with distinct first symbols: $\{0a, 1b, 2c, 3d, 4e\}$. We must choose a, b, c, d, e so that the receiver will be able to determine which codeword was transmitted. A simple repetition code will not work, since if, say, 22 is transmitted, then 11 might be received, and the receiver could not tell whether the codeword was 00 or 22. Instead, using codewords of the form (i+1 mod 5, 2i+1 mod 5) yields the code 11,23,30,42,04.

Here is the decoding table for the pentagon channel:

0 4 0 . 4 3 . 2 3 2 0 1 0 1 . 3 4 . 3 4 . 1 2 1 2

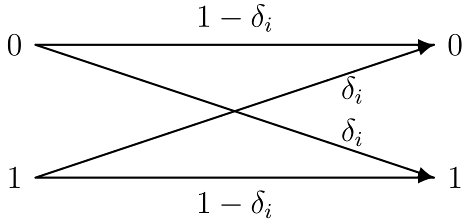
It is amusing to note that the five pairs that cannot be received are exactly the 5 codewords.

Then whenever xy is received, there is exactly one possible codeword. (Each codeword will be received as one of 4 possible 2-tuples; so there are 20 possible received 2-tuples, out of a total of 25.) Since there are 5 possible error-free messages with 2 channel uses, the zero-error capacity of this channel is at least $\frac{1}{2}\log_2 5 = 1.161$ bits.

In fact, the zero-error capacity of this channel is exactly $\frac{1}{2}\log_2 5$. This result was obtained by László Lovász, "On the Shannon capacity of a graph," *IEEE*

Transactions on Information Theory, Vol IT-25, pp. 1–7, January 1979. The first results on zero-error capacity are due to Claude E. Shannon, "The zero-error capacity of a noisy channel," *IEEE Transactions on Information Theory*, Vol IT-2, pp. 8–19, September 1956, reprinted in *Key Papers in the Development of Information Theory*, David Slepian, editor, IEEE Press, 1974.

2. **Time-varying channels.**
Consider a time-varying discrete *memoryless* channel. Let $Y_1, Y_2, \dots, Y_n$ be conditionally independent given $X_1, X_2, \dots, X_n$, with conditional distribution given by $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n p_i(y_i|x_i)$ (where $p_i(y_i|x_i)$ is a $BSC(\delta_i)$ as shown in figure). Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$.



In this problem, we show that

$$\max_{P_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^n (1 - h(\delta_i))$$

(a) Show that $I(\mathbf{X};\mathbf{Y}) \le \sum_{i=1}^n (1 - h_2(\delta_i))$ for any $P_{\mathbf{X}}$.
Hint: Use a chain of inequalities similar to the channel coding converse proof.

(b) Find a distribution over $\mathbf{X}$ for which $I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^n (1 - h_2(\delta_i))$.

Solution:

(a) We can use a similar chain of inequalities as in the proof of the converse to the channel coding theorem. Hence

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) \\ &= H(Y^n) - \sum_{i=1}^n H(Y_i|Y^{i-1}, X^n) \\ &= H(Y^n) - \sum_{i=1}^n H(Y_i|X_i) \end{aligned}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything else. Continuing the series of inequalities, we have

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\ &\leq \sum_{i=1}^{n} (1 - h_2(\delta_i)) \end{aligned}$$

The first inequality follows by chain rule + conditioning reduces entropy (equality when Y_i's are independent). The second inequality holds because Y_i is binary (equality when Y_i's are uniform in $\{0, 1\}$).

(b) We have equality when $X_1, X_2, ..., X_n$ is chosen i.i.d. $\sim Bern(1/2)$ (and so Y_i's are also i.i.d. and $Bern(1/2)$). Hence

$$\max_{p(\mathbf{x})} I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n} (1 - h_2(\delta_i))$$

3. **Suboptimal codes**.
Consider the Z channel, described by the probability transition matrix

$$p(y|x) = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

Assume that we choose a $(2^{nR}, n)$ code at random, where each codeword is a sequence of *fair* coin tosses. Find the maximum rate R such that the probability of error $P_e^{(n)}$, averaged over the randomly generated codes, tends to zero as the block length n tends to infinity.

Solution: Note that such a code will not achieve capacity, so we cannot simply claim that the maximum rate is the channel capacity. From the proof of the channel coding theorem, it follows that using a random code with codewords generated according to probability $p(x)$, we can send information at a rate $I(X; Y)$ corresponding to that $p(x)$ with an arbitrarily low probability of error. For the Z channel described in the previous problem, we can calculate $I(X; Y)$ for a uniform distribution on the input. The distribution on Y is (3/4, 1/4), and therefore

$$I(X;Y) = H(Y) - H(Y|X) = H(\frac{3}{4}, \frac{1}{4}) - \frac{1}{2} H(\frac{1}{2}, \frac{1}{2}) = \frac{3}{2} - \frac{3}{4} \log 3. \qquad (1)$$

4. **Fano's inequality.** Let $\Pr(X = i) = p_i,\ i = 1, 2, \ldots, m$ and let $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_m$. The minimal probability of error predictor of X is $\hat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$. Maximize $H(X)$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of H. This is Fano's inequality in the absence of conditioning.

Hint: Consider PMF $(p_2/P_e, p_3/P_e, \ldots, p_m/P_e)$.

Solution: (*Fano's Inequality.*) The minimal probability of error predictor when there is no information is $\hat{X} = 1$, the most probable value of X. The probability of error in this case is $P_e = 1 - p_1$. Hence if we fix P_e, we fix p_1. We maximize the entropy of X for a given P_e to obtain an upper bound on the entropy for a given P_e. The entropy,

$$\begin{aligned} H(\mathbf{p}) &= -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i && (2) \\ &= -p_1 \log p_1 - \sum_{i=2}^{m} P_e \frac{p_i}{P_e} \log \frac{p_i}{P_e} - P_e \log P_e && (3) \\ &= H(P_e) + P_e H\left(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \ldots, \frac{p_m}{P_e}\right) && (4) \\ &\le H(P_e) + P_e \log(m-1), && (5) \end{aligned}$$

since the maximum of $H\left(\frac{p_2}{P_e}, \frac{p_3}{P_e}, \ldots, \frac{p_m}{P_e}\right)$ is attained by an uniform distribution. Hence any X that can be predicted with a probability of error P_e must satisfy

$$H(X) \le H(P_e) + P_e \log(m-1), \tag{6}$$

which is the unconditional form of Fano's inequality. We can weaken this inequality to obtain an explicit lower bound for P_e,

$$P_e \ge \frac{H(X) - 1}{\log(m-1)}. \tag{7}$$

5. **Modulating Switch**
Consider the following (memoryless) channel. It has a side switch U that can be in positions ON and OFF. If U is on then the channel from X to Y is BSC_δ and if U is off then Y is Bern(1/2) regardless of X. The receiving party sees Y but not U. A design constraint is that U should be in the ON position no more than the fraction s of all channel uses, $0 \le s \le 1$.

   (a) One strategy is to put U into ON over the first sn time units and ignore the rest of the $(1-s)n$ readings of Y. What is the maximal rate in bits per channel use achievable with this strategy?

   (b) Can we increase the communication rate if the encoder is allowed to modulate the U switch together with the input X (while still satisfying the s-constraint on U)? (Hint 1: Consider the problem of communication across a channel from (U^n, X^n) to Y^n.) (Hint 2: Although the channel given in Hint 1 is not iid, Fano's inequality still applies and so the converse of the channel coding theorem still holds.)

   (c) Now assume nobody has access to U, which is iid Bern(s) independent of X. Find the capacity.

   Solution:

(a) We first note that by Fano's inequality $C^{(I)} \geq C$. Then $C^{(I)}$ can be written as

$$\begin{aligned}
C^{(I)} &= \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} I(X^n;Y^n) \\
&= \lim_{n\to\infty} \frac{1}{n} \left(\sup_{P_{X^n}} I(X^n;Y^{sn}) + I(X^n;Y_{sn+1}^n|Y^{sn}) \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \left(\sup_{P_{X^n}} I(X^n;Y^{sn}) + H(Y_{sn+1}^n \mid Y^{sn}) - H(Y_{sn+1}^n \mid X^n, Y^{sn}) \right) \\
&= \lim_{n\to\infty} \frac{1}{n} \sup_{P_{X^n}} I(X^{sn};Y^{sn}) \\
&= s(1-h_2(\delta)).
\end{aligned}$$

To achieve this capacity, we can code over BSC_δ for the first sn uses of the channel. This gives a rate of $s(1-h_2(\delta))$, as we can have a message set of size $2^{sn(1-h(\delta)-\varepsilon)}$ for any $\varepsilon > 0$.

(b) We cannot. To see this, we show that the information capacity does not increase, and use the fact that $C^{(I)} \geq C$. To see this, we note that

$$I(X^n, U^n; Y^n) = H(Y^n) - H(Y^n \mid U^n, X^n) \leq n - H(Y^n \mid U^n, X^n)$$

and we note that $H(Y^n \mid U^n, X^n) = (n - sn) + snh(\delta)$ because the sn values corresponding to when the switch is ON have entropy $snh(\delta)$ in total as they are Bern(δ) conditioned on X^n, and the other entries have a bit of entropy each. Then the information capacity is upper bounded by $s(1-h_2(\delta))$, so the Shannon capacity therefore is also $s(1-h_2(\delta))$.

(c) If nobody has access to U, then this problem is equivalent to communication over a BSC channel with parameter $s\delta + (1-s)/2$, so the capacity of this channel is $1 - h_2(s\delta + (1-s)/2)$.

6. **BSC with feedback.** Suppose that feedback is used for a binary symmetric channel with crossover probability parameter p. Each time a channel output is received, it becomes the next transmission: X_1 is Bern(1/2), $X_2 = Y_1,\ X_3 = Y_2,\ \ldots,\ X_n = Y_{n-1}$. Find $\lim_{n\to\infty} \frac{1}{n} I(X^n;Y^n)$. How does it compare to the capacity of this channel?

Solution: BSC with feedback solution.

$$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n).$$

$$H(Y^n|X^n) = \sum_{i=1}^n H(Y_i|Y^{i-1}, X^n) = H(Y_1|X_1) + \sum_{i=2}^n H(Y_i|X^n) = H(p) + 0.$$

$$H(Y^n) = \sum_{i=1}^n H(Y_i|Y^{i-1}) = H(Y_1) + \sum_{i=2}^n H(Y_i|X_i) = 1 + (n-1)H(p)$$

So,

$$I(X^n;Y^n) = 1 + (n-1)H(p) - H(p) = 1 + (n-2)H(p)$$

and,

$$\lim_{n\to\infty} \frac{1}{n} I(X^n;Y^n) = \lim_{n\to\infty} \frac{1 + (n-2)H(p)}{n} = H(p)$$

For the BSC $C = 1 - H(p)$. Depending on p, this can be either greater than or smaller than $H(p)$. For $p = 0.5$, $C = 0$, while $\lim_{n\to\infty} \frac{1}{n} I(X^n;Y^n) = H(0.5) = 1$. So the capacity of the BSC is smaller in this case. On the other hand, when $p = 0$ or 1, the capacity of the BSC is 1, but $\lim_{n\to\infty} \frac{1}{n} I(X^n;Y^n) = H(0) = 0$.