Semester A 2009/10

Solution to Practice Questions for the Final Practice

1. Initial conditions. Show, for a Markov chain, that

$$H(X_0|X_n) \ge H(X_0|X_{n-1}).$$

Thus initial conditions X_0 become more difficult to recover as the future X_n unfolds.

Solutions Initial conditions. For a Markov chain, by the data processing theorem, we have

$$I(X_0; X_{n-1}) \ge I(X_0; X_n).$$
(1)

Therefore

$$H(X_0) - H(X_0|X_{n-1}) \ge H(X_0) - H(X_0|X_n)$$
(2)

or $H(X_0|X_n)$ increases with n.

2. Entropy of a Stationary Source. Let X_1, X_2, \ldots be a discretevalued stationary random process, with entropy rate $H(\mathcal{X})$, show that for $n \ge 1$

$$H(\mathcal{X}) \le \frac{H(X^n)}{n}$$

Solutions to Entropy of a Stationary Source. We first show that $H(X_n|X^{n-1})$ is decreasing in n. Consider

$$H(X_{n+1}|X_1^n) \leq H(X_{n+1}|X_2^n)$$
 (3)

$$= H(X_n | X_1^{n-1}). (4)$$

Now we have

$$\frac{H(X^n)}{n} = \frac{1}{n} \sum_{i=1}^n H(X_i | X^{i-1})$$
(5)

$$= \frac{1}{n} \sum_{i=1}^{n} H(X_n | X_{n-i}^{n-1})$$
 (6)

$$\geq \frac{1}{n} \sum_{i=1}^{n} H(X_n | X^{n-1})$$
 (7)

$$= H(X_n|X^{n-1}) \tag{8}$$

$$\geq \lim_{n \to \infty} H(X_n | X^{n-1}) \tag{9}$$

$$= H(\mathcal{X}). \tag{10}$$

- 3. Jointly Typical Sequences. Let $A_{\epsilon}^{(n)}(X,Y)$ be the set of ϵ -strongly typical sequences (x^n, y^n) with respect to p(x, y).
 - (a) Let $x^n \in A_{\epsilon}^{(n)}$ and define

$$A_{\epsilon}^{(n)}(Y|x^n) = \left\{ y^n : (x^n, y^n) \in A_{\epsilon}^{(n)} \right\}.$$

Show that $|A_{\epsilon}^{(n)}(Y|x^n)| \doteq 2^{n(H(Y|X))}$.

(b) Consider two randomly generated codebooks, $C_1 = \{x_1^n, x_2^n, \ldots, x_{2^{nR_1}}^n\}$, where the codewords are independent and each generated according to $\sim \prod_{i=1}^n p(x_i)$, and $C_2 = \{y_1^n, y_2^n, \ldots, y_{2^{nR_1}}^n\}$, where the codewords are independent and each generated according to $\sim \prod_{i=1}^n p(x_i)$. The pmfs p(x) and p(y) are the marginals of p(x, y). Define the set

$$\mathcal{C} = \left\{ (x^n, y^n) \in \mathcal{C}_1 \times \mathcal{C}_2 \text{ such that } (x^n, y^n) \in A_{\epsilon}^{(n)}(X, Y) \right\}$$

Show that

$$E|\mathcal{C}| \doteq 2^{n\left(R_1+R_2-I(X;Y)\pm 3\epsilon\right)},$$

where the expectation is over the $C_1 \times C_2$ sets.

Solutions to Jointly Typical Sequences.

(a) For n sufficiently large,

$$\sum_{x^n \in A_{\epsilon}^{(n)}(X)} \sum_{y^n \in A_{\epsilon}^{(n)}(Y|x^n)} p(x^n, y^n) \ge \sum_{(x^n, y^n) \in A_{\epsilon}^{(n)}(X, Y)} p(x^n, y^n) \ge 1 - \epsilon.$$

Therefore,

$$\sum_{x^n \in A_{\epsilon}^{(n)}(X)} p(x^n) \sum_{y^n \in A_{\epsilon}^{(n)}(Y|x^n)} p(y^n|x^n) \ge 1 - \epsilon.$$
(11)

Since $p(y^n|x^n) \leq 2^{-n(H(Y|X)-2\epsilon)}$, we have

$$\sum_{x^n \in A_{\epsilon}^{(n)}(X)} p(x^n) \sum_{y^n \in A_{\epsilon}^{(n)}(Y|x^n)} 2^{-n(H(Y|X)-2\epsilon)} \ge 1 - \epsilon, \qquad (12)$$

or

$$\sum_{x^n \in A_{\epsilon}^{(n)}(X)} p(x^n) |A_{\epsilon}^{(n)}(Y|x^n)| \ge (1-\epsilon) 2^{n(H(Y|X)-2\epsilon)}.$$
 (13)

Hence,

$$E_{X^n}|A_{\epsilon}^{(n)}(Y|x^n)| \ge (1-\epsilon)2^{n(H(Y|X)-2\epsilon)}.$$
 (14)

(b) Let

$$i(x^{n}, y^{n}) = \begin{cases} 1 & \text{if } (x^{n}, y^{n}) \in A_{\epsilon}^{(n)}(X, Y), \\ 0 & \text{if } (x^{n}, y^{n}) \notin A_{\epsilon}^{(n)}(X, Y). \end{cases}$$
(15)

Then,

$$|\mathcal{C}| = \sum_{(X^n, Y^n) \in \mathcal{C}_1 \times \mathcal{C}_2} i(X^n, Y^n).$$
(16)

End,

$$E|\mathcal{C}| = |\mathcal{C}_1 \times \mathcal{C}_2| P\{i(X^n, Y^n)\}$$

$$(17)$$

$$e^{n(B_1 + B_2)} P\{i(X^n, Y^n)\}$$

$$(17)$$

$$= 2^{n(R_1+R_2)} P\{i(X^n, Y^n)\}$$
(18)

$$\doteq 2^{n(R_1 + R_2 - I(X;Y) \pm 3\epsilon)}, \tag{19}$$

where the expectation is taken over the choice of codebooks.

4. Memoryless channel without feedback In the lectures we stated that for a DMC with no feedback the definition of memory-lessness; $p(y_i|x^i, y^{i1}) = p(y_i|x_i), i = 1, 2, ..., n$, implies that $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$. Use the fact that with no feedback $W \to X^n \to Y^n$ to prove this claim.

Solutions to Memoryless channel without feedback We use the definition of memorylessness given in the lecture. Consider

$$p(y^{n}|x^{n}) = \prod_{i=1}^{n} p(y_{i}|x^{n}, y^{i-1})$$
(20)

$$= \prod_{i=1}^{n} p(y_i | x^i, x_{i+1}^n, y^{i-1})$$
(21)

$$= \prod_{i=1}^{n} \frac{p(x_{i+1}^{n}|y^{i}, x^{i})p(x^{i}, y^{i})}{p(x^{i}, x_{i+1}^{n}, y^{i-1})}$$
(22)

$$= \prod_{i=1}^{n} \frac{p(x_{i+1}^{n}|y^{i}, x^{i})p(y_{i}|x^{i}, y^{i-1})p(x^{i}, y^{i-1})}{p(x_{i+1}^{n}|x^{i}, y^{i-1})p(x^{i}, y^{i-1})}$$
(23)

$$= \prod_{i=1}^{n} \frac{p(x_{i+1}^{n} | x^{i}) p(y_{i} | x^{i}, y^{i-1})}{p(x_{i+1}^{n} | x^{i})}$$
(24)

$$= \prod_{i=1}^{n} p(y_i | x^i, y^{i-1}) = \prod_{i=1}^{n} p(y_i | x_i), \qquad (25)$$

where $p(x_{i+1}^{n}|x^{i}, y^{i}) = p(x_{i+1}^{n}|x^{i}, y^{i-1}) = p(x_{i+1}^{n}|x^{i})$, since there is no feedback.

5. Maximum entropy process. A discrete memoryless source has alphabet $\{1, 2\}$ where the symbol 1 has duration 1 and the symbol 2 has duration 2. The probabilities of 1 and 2 are p_1 and p_2 , respectively. Find the value of p_1 that maximizes the source entropy per unit time $H(X)/El_X$. What is the maximum value H?

Solutions Maximum entropy process. The entropy per symbol of the source is

$$H(p_1) = -p_1 \log p_1 - (1 - p_1) \log(1 - p_1)$$

and the average symbol duration (or time per symbol) is

$$T(p_1) = 1 \cdot p_1 + 2 \cdot p_2 = p_1 + 2(1 - p_1) = 2 - p_1 = 1 + p_2.$$

Therefore the source entropy per unit time is

$$f(p_1) = \frac{H(p_1)}{T(p_1)} = \frac{-p_1 \log p_1 - (1 - p_1) \log(1 - p_1)}{2 - p_1}$$

Since f(0) = f(1) = 0, the maximum value of $f(p_1)$ must occur for some point p_1 such that $0 < p_1 < 1$ and $\partial f / \partial p_1 = 0$ and

$$\frac{\partial}{\partial p_1} \frac{H(p_1)}{T(p_1)} = \frac{T(\partial H/\partial p_1) - H(\partial T/\partial p_1)}{T^2}$$

After some calculus, we find that the numerator of the above expression (assuming natural logarithms) is

$$T(\partial H/\partial p_1) - H(\partial T/\partial p_1) = \ln(1-p_1) - 2\ln p_1$$

which is zero when $1 - p_1 = p_1^2 = p_2$, that is, $p_1 = \frac{1}{2}(\sqrt{5} - 1) = 0.61803$, the reciprocal of the golden ratio, $\frac{1}{2}(\sqrt{5} + 1) = 1.61803$. The corresponding entropy per unit time is

$$\frac{H(p_1)}{T(p_1)} = \frac{-p_1 \log p_1 - p_1^2 \log p_1^2}{2 - p_1} = \frac{-(1 + p_1^2) \log p_1}{1 + p_1^2} = -\log p_1 = 0.69424 \text{ bits.}$$

Note that this result is the same as the maximum entropy rate for the Markov chain in problem 4.7(d). This is because a source in which every 1 must be followed by a 0 is equivalent to a source in which the symbol 1 has duration 2 and the symbol 0 has duration 1.

6. Horse race. Consider a horse race with 4 horses. Assume that each of the horses pays 4-for-1 if it wins. Let the probabilities of winning of the horses be $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$. If you started with \$100 and bet optimally to maximize your long term growth rate, what are your optimal bets on each horse? Approximately how much money would you have after 20 races with this strategy ?

Horse race. The optimal betting strategy is proportional betting, i.e., dividing the investment in proportion to the probabilities of each horse winning. Thus the bets on each horse should be (50%, 25%, 12.5%, 12.5%), and the growth rate achieved by this strategy is equal to $\log 4 - H(p) = \log 4 - H(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}) = 2 - 1.75 = 0.25$. After 20 races with this strategy, the wealth is approximately $2^{nW} = 2^5 = 32$, and hence the wealth would grow approximately 32 fold over 20 races.

7. Weak Typicality vs. Strong Typicality

In this problem, we compare the weakly typical set $A_{\epsilon}(P)$ and the strongly typical set $T_{\delta}(P)$. To recall, the definition of two sets are following.

$$A_{\epsilon}(P) = \left\{ x^{n} \in \mathcal{X}^{n} : \left| -\frac{1}{n} \log P^{n}(x^{n}) - H(P) \right| \le \epsilon \right\}$$
$$T_{\delta}(P) = \left\{ x^{n} \in \mathcal{X}^{n} : \|P_{x^{n}} - P\|_{\infty} \le \frac{\delta}{|\mathcal{X}|} \right\}$$

- (a) Suppose P is such that P(a) > 0 for all $a \in \mathcal{X}$. Then, there is an inclusion relationship between the weakly typical set $A_{\epsilon}(P)$ and the strongly typical set $T_{\delta}(P)$ for an appropriate choice of ϵ . Which of the statement is true: $A_{\epsilon}(P) \subseteq T_{\delta}(P)$ or $A_{\epsilon}(P) \supseteq$ $T_{\delta}(P)$? What is the appropriate relation between δ and ϵ ?
- (b) Give a description of the sequences that belongs to $A_{\epsilon}(P)$, vs. the sequences that belongs to $T_{\delta}(P)$, when the source is uniformly distributed, i.e. $P(a) = \frac{1}{|\mathcal{X}|}, \forall a \in \mathcal{X}$. (Assume $|\mathcal{X}| < \infty$.)
- (c) Can you explain why $T_{\delta}(P)$ is called **strongly** typical set and $A_{\epsilon}(P)$ is called **weakly** typical set?

Solution: Weak Typicality vs. Strong Typicality

(a) From Problem 2, we can see that if $x^n \in T_{\delta}(P)$, then

$$\left| -\frac{1}{n} \log P^n(x^n) - H(P) \right| \le \delta' = \frac{\delta}{|\mathcal{X}|} \sum_a \log \frac{1}{P(a)}$$

Therefore, we can see that if $\epsilon > \frac{\delta}{|\mathcal{X}|} \sum_{a} \log \frac{1}{P(a)}$, then $A_{\epsilon}(P) \supseteq T_{\delta}(P)$. To show that the other way does not hold, see the next part.

(b) When P is a uniform distribution, we can see that $A_{\epsilon}(P)$ includes every possible sequences x^n . To see this, we can easily see that $P^n(x^n) = (\frac{1}{|\mathcal{X}|})^n$, and thus, $-1/n \log P^n(x^n) = \log |\mathcal{X}| = H(P)$. Thus, $x^n \in A_{\epsilon}(P)$ for all x^n , for all $\epsilon > 0$. However, obviously, among those sequences in $A_{\epsilon}(P)$, only those who have the type

$$\frac{1}{|\mathcal{X}|} - \frac{\delta}{|\mathcal{X}|} \le P_{x^n}(a) \le \frac{1}{|\mathcal{X}|} + \frac{\delta}{|\mathcal{X}|}$$

for each letter $a \in \mathcal{X}$, are in $T_{\delta}(P)$. Therefore, for sufficiently small value of δ , there always exist some sequences that are in $A_{\epsilon}(P)$, but not in $T_{\delta}(P)$.

- (c) From above questions, we can see that the strongly typical set is contained in the weakly typical set for appropriate choice of δ and ϵ . Thus, we can see that the definition of strong typical set is stronger than that of the weakly typical set.
- 8. Sanov's theorem: Prove the simple version of Sanov's theorem for the binary random variables, i.e., let X_1, X_2, \ldots, X_n be a sequence of binary random variables, drawn i.i.d. according to the distribution:

$$\Pr(X=1) = q, \quad \Pr(X=0) = 1 - q.$$
 (26)

Let the proportion of 1's in the sequence X_1, X_2, \ldots, X_n be $p_{\mathbf{X}}$, i.e.,

$$p_{X^n} = \frac{1}{n} \sum_{i=1}^n X_i.$$
 (27)

By the law of large numbers, we would expect $p_{\mathbf{X}}$ to be close to q for large n. Sanov's theorem deals with the probability that p_{X^n} is far away from q. In particular, for concreteness, if we take $p > q > \frac{1}{2}$, Sanov's theorem states that

$$-\frac{1}{n}\log\Pr\left\{(X_1, X_2, \dots, X_n) : p_{X^n} \ge p\right\} \to p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q} = D((p, 1-p)||(q, 1-q)|)$$
(28)

Justify the following steps:

•

$$\Pr\{(X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p\} \le \sum_{i=\lfloor np \rfloor}^n \binom{n}{i} q^i (1-q)^{n-i} \quad (29)$$

- Argue that the term corresponding to $i = \lfloor np \rfloor$ is the largest term in the sum on the right hand side of the last equation.
- Show that this term is approximately 2^{-nD} .
- Prove an upper bound on the probability in Sanov's theorem using the above steps. Use similar arguments to prove a lower bound and complete the proof of Sanov's theorem.

Solutions to Sanov's theorem

• Since $n\overline{X}_n$ has a binomial distribution, we have

$$\Pr(n\overline{X}_n = i) = \binom{n}{i} q^i (1-q)^{n-i}$$
(30)

and therefore

$$\Pr\{(X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p\} \le \sum_{i=\lfloor np \rfloor}^n \binom{n}{i} q^i (1-q)^{n-i} \quad (31)$$

•

$$\frac{\Pr(n\overline{X}_n = i+1)}{\Pr(n\overline{X}_n = i)} = \frac{\binom{n}{i+1}q^{i+1}(1-q)^{n-i-1}}{\binom{n}{i}q^i(1-q)^{n-i}} = \frac{n-i}{i+1}\frac{q}{1-q} \quad (32)$$

This ratio is less than 1 if $\frac{n-i}{i+1} < \frac{1-q}{q}$, i.e., if i > nq - (1-q). Thus the maximum of the terms occurs when $i = \lfloor np \rfloor$.

• From Example 11.1.3,

$$\binom{n}{\lfloor np \rfloor} \doteq 2^{nH(p)} \tag{33}$$

and hence the largest term in the sum is

$$\binom{n}{\lfloor np \rfloor} q^{\lfloor np \rfloor} (1-q)^{n-\lfloor np \rfloor} = 2^{n(-p\log p - (1-p)\log(1-p)) + np\log q + n(1-p)\log(1-q)} = 2^{-nD(p||q)}$$
(34)

• From the above results, it follows that

$$\Pr\left\{ (X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p \right\} \le \sum_{i=\lfloor np \rfloor}^n \binom{n}{i} q^i (1-q)^{n-i} \quad (35)$$
$$\le (n - \lfloor np \rfloor) \binom{n}{\lfloor np \rfloor} q^i (1-q)^{36} q^i (1$$

where the second inequality follows from the fact that the sum is less than the largest term times the number of terms. Taking the logarithm and dividing by n and taking the limit as $n \to \infty$, we obtain

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr\{(X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p\} \le -D(p||q)$$
(38)

Similarly, using the fact the sum of the terms is larger than the largest term, we obtain

$$\Pr\left\{ (X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p \right\} \ge \sum_{i=\lceil np \rceil}^n \binom{n}{i} q^i (1-q)^{n-i} (39)$$
$$\ge \binom{n}{\lceil np \rceil} q^i (1-q)^{n-i} (40)$$

$$\geq 2^{-nD(p||q)} \tag{41}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr\left\{ (X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p \right\} \ge -D(p||q)$$
(42)

Combining these two results, we obtain the special case of Sanov's theorem

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left\{ (X_1, X_2, \dots, X_n) : p_{\mathbf{X}} \ge p \right\} = -D(p||q)$$
(43)

9. successive cancellation.

Consider a DM-MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$. To achieve a corner point of a set $\mathcal{R}(X_1, X_2)$, e.g., $R_1 = I(X_1; Y\mathcal{X}_2) + \epsilon$, $R_2 = I(X_2; Y) + \epsilon$ for any $\epsilon > 0$, use random coding and the following two-step decoding scheme: the receiver first declares that \hat{w}_2 is sent if it is the unique message such that $((x_2^n(\hat{w}_2), y^n) \in A_{\epsilon}^{(n)})$, otherwise, an error is declared, if such \hat{w}_2 is found, the receiver declares that \hat{w}_1 is sent if the unique message such that $((x_1^n(\hat{w}_1), x_2^n(\hat{w}_2), y^n) \in A_{\epsilon}^{(n)})$, otherwise an an error is declared. Provide detailed analysis of error probability to show that this corner point is achievable.

Solutions of successive cancellation. Without loss of generality, we can assume that the transmitted indices are $W_1, W_2 = (1, 1)$. Define,

$$E_{2j} = \{ (x_2^n(j), y^n) \in A_{\epsilon}^n(X_2, Y) \}$$

$$E_{1i} = \{ (x_1^n(i), x_2^n(1), y^n) \in A_{\epsilon}^n(X_1, X_2, Y) \}.$$
(44)

Therefore,

$$\begin{aligned} P_e^{(n)} &= P(E_{21}^c \bigcup \cup_{j \neq 1} E_{2j} \bigcup \cup_{i \neq 1} E_{1i}) \\ &\leq P(E_{21}^c) + \sum_{j \neq 1} P(E_{2j}) + \sum_{i \neq 1} P(E_{1i}) \\ &\leq \epsilon + \sum_{j \neq 1} 2^{-n(I(X_2;Y) - \epsilon)} + \sum_{i \neq 1} 2^{-n(I(X_1;Y|X_2) - \epsilon)} \\ &\leq \epsilon + 2^{n(R_2 - I(X_2;Y) + \epsilon)} + 2^{n(R_1 - I(X_1;Y|X_2) + \epsilon)}, \end{aligned}$$

and probability of error approaches zero if

$$R_1 < I(X_1; Y | X_2) - \epsilon$$

$$R_2 < I(X_2; Y) - \epsilon$$
(45)

10. Cooperative Capacity of a MAC.

Consider a DM-MAC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$. Assume that both senders have access to both messages $W_1 \in \{1, 2, ...2^{nR}\}$ and $W_2 \in \{1, 2, ...2^{nR}\}$, thus the codewords $X_1(W_1, W_2)$ and $X_2(W_1, W_2)$ can depend on both messages.

- (a) Find the capacity region.
- (b) Evaluate the region for the AWGN MAC with noise power N and power constraints P_1 and P_2

Solutions of Cooperative Capacity of a MAC.

(a) Since both transmitters have access to both messages, we only need to bound the summation of rates, because the rate of each individual message can be as high as the maximum rate. The capacity region is given by

$$R_1 + R_2 \le \max_{p(x_1, x_2)} I(X_1, X_2; Y)$$
(46)

(b) Consider

$$R_{1} + R_{2} \leq I(X_{1}, X_{2}; Y)$$

$$= h(Y) - h(Y|X_{1}, X_{2})$$

$$= h(X_{1} + X_{2} + Z) - h(Z)$$

$$\leq \frac{1}{2}\log(2\pi e(E(X_{1} + X_{2})^{2} + N)) - \frac{1}{2}\log(2\pi eN)$$

$$= \frac{1}{2}\log\left(1 + \frac{E(X_{1} + X_{2})^{2}}{N}\right)$$

$$= \frac{1}{2}\log\left(1 + \frac{E(X_{1})^{2} + E(X_{2})^{2} + 2E(X_{1}X_{2})}{N}\right)$$

$$\leq \frac{1}{2}\log\left(1 + \frac{P_{1} + P_{2} + 2\sqrt{P_{1}P_{2}}}{N}\right)$$

$$= C\left(\frac{P_{1} + P_{2} + 2\sqrt{P_{1}P_{2}}}{N}\right)$$

11. Time-sharing for MAC In the examples in the class (including the AWGN case) can all be expressed as union of $\mathcal{R}(X_1, X_2)$ sets and no time-sharing is necessary. Is time-sharing ever necessary? The answer is YES. Find the capacity of the *push to talk* MAC channel with binary inputs and output and p(0|0,0) = p(1|0,1) = p(1|1,0) = 1 and p(0|1,1) = 1/2. Why is this channel called "push to talk"? Show that the capacity region cannot be completely expressed as the union of $\mathcal{R}(X_1, X_2)$ sets and that time-sharing (convexification) is necessary.

Solutions of Time-sharing for MAC

Suppose $X_1 \sim Ber(p_1)$ and $X_2 \sim Ber(p_2)$. It can be easily shown that $Y \sim Ber(p_1 + p_2 - \frac{3p_1p_2}{2})$ and

$$I(X_1, X_2; Y) = H\left(p_1 + p_2 - \frac{3p_1p_2}{2}\right) - p_1p_2.$$
 (47)

IF $p_1p_2 \neq 0$, we have $I(X_1, X_2; Y) < 1$. However, the points $(R_1, R_1) = (1, 0)$ and $(R_1, R_2) = (0, 1)$ are achievable if one of the transmitters fixes its output at zero, i.e., $p_1 = 0$ or $p_2 = 0$. Therefore, by time sharing, all the points on the line $R_1 + R_2 = 1$ are achievable, however, these points do not belong to any region of the form $\mathcal{R}(X_1, X_2)$

12. MAC capacity with costs

The cost of using symbol x is r(x). The cost of a codeword x^n is $r(x^n) = \frac{1}{n} \sum_{i=1}^n r(x_i)$. A $(2^{nR}, n)$ codebook satisfies cost constraint r if $\frac{1}{n} \sum_{i=1}^n r(x_i(w)) \leq r$, for all $w \in 2^{nR}$.

- (a) Find an expression for the capacity C(r) of a discrete memoryless channel with cost constraint r.
- (b) Find an expression for the multiple access channel capacity region for $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$ if sender X_1 has cost constraint r_1 and sender X_2 has cost constraint r_2 .
- (c) Prove the converse for (b).

Solutions of MAC capacity with costs

(a) The capacity of a discrete memoryless channel with cost constraint r is given by

$$C(r) = \max_{p(x):\sum_{x} p(x)r(x) \le r} I(X;Y).$$

$$(48)$$

The achievability follows immediately from Shannon's 'average over random codebooks' method and joint typicality decoding. (See Section 9.1 for the power constraint example.)

For the converse, we need to establish following simple properties of the capacity-cost function C(r).

Theorem The capacity cost function C(r) given in (48) is a nondecreasing concave function of r. % bftheorem

Remark: These properties of the capacity cost function C(r) exactly parallel those of the rate distortion function R(D). (See Lemma 10.4.1 of the text.)

Proof: The monotonicity is a direct consequence of the definition of C(r). To prove the concavity, consider two points (C_1, r_1) and (C_2, r_2) which lie on the capacity cost curve. Let the distributions that achieve these pairs be $p_1(x)$ and $p_2(x)$. Consider the distribution $p_{\lambda} = \lambda p_1 + (1 - \lambda)p_2$. Since the cost is a linear function of the distribution, we have $r(p_{\lambda}) = \lambda r_1 + (1 - \lambda)r_2$. Mutual information, on the other hand, is a concave function of the input distribution (Theorem 2.7.4) and hence

$$C(\lambda r_1 + (1 - \lambda)r_2) = C(r(p_\lambda))$$
(49)

$$\geq I_{p_{\lambda}}(X;Y) \tag{50}$$

$$\geq \lambda I_{p_1}(X;Y) + (1-\lambda)I_{p_2}(X;Y)$$
 (51)

$$= \lambda C(r_1) + (1 - \lambda)C(r_2), \qquad (52)$$

which proves that C(r) is concave in r. \Box Now we are ready to prove the converse. Consider any $(2^{nR}, n)$ code that satisfies the cost constraint

$$\frac{1}{n}\sum_{i=1}^{n}r(x_i(w)) \le r$$

for $w = 1, 2, \ldots, 2^{nR}$, which in turn implies that

$$\frac{1}{n}\sum_{i=1}^{n}E(r(X_i)) \le r,$$
(53)

where the expectation is with respect to the uniformly drawn message index W. As in the case without the cost constraint, we begin with Fano's inequality to obtain the following chain of inequalities:

$$nR = H(W) \tag{54}$$

$$\leq I(W; Y^{n}) + n\epsilon_{n}$$

$$\leq I(X^{n}; Y^{n}) + n\epsilon_{n}$$
(55)
(56)

$$\leq I(X^n; Y^n) + n\epsilon_n \tag{56}$$

$$\leq H(Y^n) - H(Y^n|X^n) + n\epsilon_n \tag{57}$$

$$\leq \sum_{i=1} H(Y_i) - H(Y_i|X^n, Y^{i-1}) + n\epsilon_n \tag{58}$$

$$= \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) + n\epsilon_n$$
(59)

$$= \sum_{i=1}^{n} I(X_i; Y_i) + n\epsilon_n \tag{60}$$

$$\stackrel{(a)}{\leq} \sum_{i=1}^{n} C(E(r(X_i))) + n\epsilon_n \tag{61}$$

$$= n \sum_{i=1}^{n} \frac{1}{n} C(E(r(X_i))) + n\epsilon_n$$
(62)

$$\stackrel{(b)}{\leq} nC\left(\frac{1}{n}\sum_{i=1}^{n}E(r(X_i))\right) + n\epsilon_n \tag{63}$$

$$\stackrel{(c)}{\leq} nC(r) + n\epsilon_n, \tag{64}$$

where

(a) follows from the definition of the capacity cost function,

(b) from the concavity of the capacity cost function and Jensen's inequality, and

(c) from Eq. (53) and the fact that C(r) is non-decreasing in r. Note that we cannot jump from (60) to (64) since $E(r(X_i))$ may be greater than r for some i.

(b) The capacity region under cost constraints r_1 and r_2 is given by the closure of the set of all (R_1, R_2) pairs satisfying

$$R_1 < I(X_1; Y | X_2, Q),$$

$$R_2 < I(X_2; Y | X_1, Q),$$

$$R_1 + R_2 < I(X_1, X_2; Y | Q)$$

for some choice of the joint distribution $p(q)p(x_1|q)p(x_2|q)p(y|x_1,x_2)$ with

$$\sum_{x_1} p(x_1)r_1(x_1) \le r_1,$$
$$\sum_{x_2} p(x_2)r_2(x_2) \le r_2,$$

and $|\mathcal{Q}| \leq 4$.

(c) Again the achievability proof is an easy extension from the case without cost constraints. For the converse, consider any sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $P_e^{(n)} \to 0$ satisfying

$$\frac{1}{n} \sum_{i} r_1(x_{1i}(w_{1i})) \leq r_1$$

$$\frac{1}{n} \sum_{i} r_2(x_{2i}(w_{2i})) \leq r_2,$$

for all $w_{1i} = 1, 2, ..., 2^{nR_1}, w_{2i} = 1, 2, ..., 2^{nR_2}$. By taking expectation with respect to the random message index pair (W_1, W_2) , we get

$$\frac{1}{n}\sum_{i} E(r_1(X_{1i})) \le r_1 \quad \text{and} \quad \frac{1}{n}\sum_{i} E(r_2(X_{2i})) \le r_2.$$
(65)

By starting from Fano's inequality and taking the exact same steps as in the converse proof for the MAC without constraints (see Section 14.3.4 of the text), we obtain

$$nR_{1} \leq \sum_{i=1}^{n} I(X_{1i}; Y_{i}|X_{2i}) + n\epsilon_{1n} = nI(X_{1Q}; Y_{Q}|X_{2Q}, Q) + n\epsilon_{1n},$$

$$nR_{2} \leq \sum_{i=1}^{n} I(X_{2i}; Y_{i}|X_{1i}) + n\epsilon_{2n} = nI(X_{2Q}; Y_{Q}|X_{1Q}, Q) + n\epsilon_{2n},$$

$$n(R_{1} + R_{2}) \leq \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{i}) + n\epsilon_{n} = nI(X_{1Q}, X_{2Q}; Y_{Q}|Q) + n\epsilon_{n},$$

where the random variable Q is uniform over $\{1, 2, ..., n\}$ and independent of (X_{1i}, X_{2i}, Y_i) for all i.

Now define $X_1 \stackrel{\triangle}{=} X_{1Q}$, $X_2 \stackrel{\triangle}{=} X_Q$, and $Y \stackrel{\triangle}{=} Y_Q$. It is easy to check that (Q, X_1, X_2, Y) have a joint distribution of the form $p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)$. Moreover, from Eq. (65),

$$\sum_{x_1} \Pr(X_1 = x_1) r_1(x_1) = \sum_{x_1} \Pr(X_{1Q} = x_1) r_1(x_1)$$

$$= \sum_{x_1} \sum_{i=1}^n \Pr(X_{1Q} = x_1 | Q = i) \Pr(Q = i) r_1(x_1)$$

$$= \sum_{x_1} \sum_{i=1}^n \frac{1}{n} \Pr(X_{1i} = x_1) r_1(x_1)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{x_1} \Pr(X_{1i} = x_1) r_1(x_1)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{x_1} \Pr(X_{1i} = x_1) r_1(x_1)$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{x_1} \Pr(X_{1i} = x_1) r_1(x_1)$$

$$= \frac{1}{n} \sum_{i=1}^n E(r_1(X_{1i}))$$

$$\leq r_1,$$

and similarly,

$$\sum_{x_2} \Pr(X_2 = x_2) r_2(x_2) \le r_2.$$

Therefore, we have shown that any sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes satisfying cost constraints with $P_e^{(n)} \to 0$ should have the rates satisfying

$$R_1 < I(X_1; Y | X_2, Q),$$

$$R_2 < I(X_2; Y | X_1, Q),$$

$$R_1 + R_2 < I(X_1, X_2; Y | Q)$$

for some choice of the joint distribution $p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)$ with

$$\sum_{x_1} p(x_1)r_1(x_1) \le r_1$$

 $\sum_{x_2} p(x_2) r_2(x_2) \le r_2.$

Finally, from Theorem 14.3.4, the region is unchanged if we limit the cardinality of Q to 4, which completes the proof of the converse.

Note that, compared to the single user case in part (a), the converse for the MAC with cost constraints is rather straightforward. Here the time sharing random variable Q saves the trouble of dealing with costs at each time index i.

and