Solution to Practice Questions for the Final Practice

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

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

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

2. **Entropy of a Stationary Source**. Let $X_1, X_2, \dots$ be a discrete-valued stationary random process, with entropy rate $H(\mathcal{X})$, show that for $n \geq 1$

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

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

   (b) Consider two randomly generated codebooks, $\mathcal{C}_1 = \{x_1^n, x_2^n, \dots, x_{2^{nR_1}}^n\}$, where the codewords are independent and each generated according to $\sim \Pi_{i=1}^n p(x_i)$, and $\mathcal{C}_2 = \{y_1^n, y_2^n, \dots, y_{2^{nR_1}}^n\}$, where the codewords are independent and each generated according to $\sim \Pi_{i=1}^n p(x_i)$. 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 $\mathcal{C}_1 \times \mathcal{C}_2$ sets.

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

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?

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?

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:

$$\Pr(X = 1) = q, \quad \Pr(X = 0) = 1 - q. \tag{1}$$

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

$$p_{X^n} = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{2}$$

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

$$-\frac{1}{n} \log \Pr\left\{(X_1, X_2, \ldots, X_n) : p_{X^n} \geq p\right\} \to p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} = D((p, 1-p)||(q, 1-q) \tag{3}$$

Justify the following steps:

- $$\Pr\left\{(X_1, X_2, \ldots, X_n) : p_{\mathbf{X}} \geq p\right\} \leq \sum_{i=\lfloor np \rfloor}^{n} \binom{n}{i} q^i (1-q)^{n-i} \tag{4}$$

- 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.

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.

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

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.

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).