Tutorial IX

Midterm preparation

Problem 1 (Basics). For each one of these statements, say whether it is true or false and provide a brief justification.

1. Define the distribution \( P_X = (1/5, 1/5, 1/5, 2/5) \). We have \( H(X) = \log_2 5 \).

2. For any random variable \( X \in \mathcal{X} \) and any \( x \in \mathcal{X} \), we have \( P_X(x) \leq 2^{-H(X)} \).

3. Define the channel \( W \) with binary input and output given by \( W(0|0) = 1/3 \), \( W(1|0) = 2/3 \), \( W(0|1) = 1/3 \), \( W(1|1) = 2/3 \). The capacity of this channel is 0.

4. Define the tripartite mutual information \( I(X : Y : Z) := I(X : Y) - I(X : Y|Z) \). For any random variables \( X, Y, Z \), we have \( I(X : Y : Z) \geq 0 \).

5. For any random variables \( X_1, X_2 \), we have \( H(X_1X_2) = H(X_1) + H(X_2) \).

6. Consider the distribution \( P_X = (1/2, 1/4, 1/8, 1/16, 1/16) \). The code with the shortest expected length for this source has expected length exactly \( H(X) \).

7. Let \( X_1, \ldots, X_n \) be iid random variables each living in the finite set \( \mathcal{X} \). A sequence \( x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n \) is said to be \( \epsilon \)-typical if \( 2^{-nH(X_1)+\epsilon} \leq P_{X_1\ldots X_n}(x_1 \ldots x_n) \leq 2^{-nH(X_1)-\epsilon} \). Now a sequence \( x^n = (x_1, \ldots, x_n) \) is said to be \( \epsilon \)-strongly typical if \( (1-\epsilon)P_{X_1}(a) \leq N(a|x^n)/n \leq (1+\epsilon)P_{X_1}(a) \) for all \( a \in \mathcal{X} \). Here \( N(a|x^n) \) denotes the number of times the symbol \( a \) occurs in the sequence \( x^n \).

The statement is that if \( x^n \) is \( \epsilon \)-strongly typical, then \( x^n \) is \( c \cdot \epsilon \)-typical where \( c \) is a constant that is independent of \( n \) but can depend on the distribution \( P_{X_1} \).

8. If \( x^n \) is \( \epsilon \)-typical, then it is also \( c \cdot \epsilon \)-strongly typical for a constant \( c \) that is independent of \( n \) but can depend on the distribution \( P_{X_1} \).

Problem 2 (Capacity of a simple channel). Define the channel \( W \) with binary input \( \mathcal{X} = \{0, 1\} \) and binary output \( \mathcal{Y} = \{0, 1\} \) and \( W(0|0) = 1 \), \( W(0|1) = 1/2 \) and \( W(1|1) = 1/2 \). Show that the information capacity \( C(W) = \sup_{x \in [0,1/2]} h_2(x) - 2x \), where \( h_2(x) = -x \log_2 x - (1-x) \log_2(1-x) \) is the binary entropy function.

Problem 3 (Compression with side information). In class, we showed that in order to compress a source \( X \in \mathcal{X} \) with distribution \( P_X \) into \( \ell \) bits, the minimum error probability \( \delta_{\text{opt}}(P_X, \ell) \) satisfies for any \( \sigma > 0 \),

\[
P \left\{ \log_2 \frac{1}{P_X(X)} > \ell + \sigma \right\} - 2^{-\sigma} \leq \delta_{\text{opt}}(P_X, \ell) \leq P \left\{ \log_2 \frac{1}{P_X(X)} > \ell \right\} .
\]

[Added remark: We did not do it this year, but in the tutorial, you proved something very similar]
As a consequence, we showed that in the case where the source $X^n$ is $n$ independent copies $X_1, \ldots, X_n$ of $X$, then

$$\lim_{n \to \infty} \delta_{\text{opt}}(P_{X^n}, Rn) = \begin{cases} 1 & \text{if } R < H(X) \\ 0 & \text{if } R > H(X) \end{cases}.$$  

In this problem, we consider variants of fixed-length compression with side information, i.e., there is a random variable $Y \in \mathcal{Y}$ correlated with the source $X$ that can be used when compressing $X$. As usual, we write $P_{XY}$ for the joint distribution of $X$ and $Y$ and this distribution is assumed to be known to everybody. Recall that we also write $P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)}$.

1. In this question, the compressor and the decompressor have access to the random variable $Y$. More precisely, a compressor is now $C : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}^\ell$ and the decompressor is a function $D : \{0, 1\}^\ell \times \mathcal{Y} \to \mathcal{X}$. The error probability is defined as $P \{ D(C(X,Y), Y) \neq X \}$. Note that the probability is over $X$ and $Y$. Let us call $\delta_{\text{opt}}(X|Y, \ell)$ the smallest error probability over all compressor-decompressor pair.

   (a) Suppose $X = Y$ with probability 1, what can you say on $\delta_{\text{opt}}(X|Y, \ell)$?
   (b) Show that $\delta_{\text{opt}}(X|Y, \ell) = \mathbb{E}_{y \sim P_Y} \{ \delta_{\text{opt}}(P_{X|Y=y}, \ell) \}$
   (c) Using Eq. (11) as a black-box, deduce that we have
   $$P \left\{ \log_2 \frac{1}{P_{X|Y}(X|Y)} > \ell + \tau \right\} - 2^{-\tau} \leq \delta_{\text{opt}}(X|Y, \ell) \leq P \left\{ \log_2 \frac{1}{P_{X|Y}(X|Y)} > \ell \right\}. $$
   (d) If we now take $n$ independent pairs $(X_i, Y_i)$ distributed according to $P_{XY}$, and let $X^n = X_1 \ldots X_n$ and $Y^n = Y_1, \ldots, Y_n$. What can you say on the limit $\lim_{n \to \infty} \delta_{\text{opt}}(X^n|Y^n, Rn)$ for different values of $R$?

2. Now we consider a setting where the compressor does not have access to $Y$. Only the decompressor sees $Y$. So the compressor is now $C : \mathcal{X} \to \{0, 1\}^\ell$ and $D : \{0, 1\}^\ell \times \mathcal{Y} \to \mathcal{X}$. The error probability is given by $P \{ D(C(X), Y) \neq X \}$. We call $\delta_{\text{opt}}(X|Y, \ell)$ the smallest error probability for such a compressor-decompressor pair in this setting.

   (a) Using the previous questions, show that
   $$P \left\{ \log_2 \frac{1}{P_{X|Y}(X|Y)} > \ell + \tau \right\} - 2^{-\tau} \leq \delta_{\text{opt}}(X|Y, \ell).$$
   (b) We choose the compressor as follows. For every $x \in \mathcal{X}$, let $B_x$ be uniformly random and independent bitstrings of length $\ell$. We set $C(x) = B_x$ for all $x \in \mathcal{X}$. Then define
   $$D(w, y) = \begin{cases} x & \text{if } x \text{ is the unique such that } C(x) = w \text{ and } \log_2 \frac{1}{P_{X|Y}(x|y)} \leq \ell - \tau \\ x_0 & \text{otherwise,} \end{cases}$$
   for some arbitrary $x_0 \in \mathcal{X}$. Show that in expectation over the choice of $B_x$ for $x \in \mathcal{X}$, the error probability of the pair $(C, D)$ is bounded above by
   $$P \left\{ \log_2 \frac{1}{P_{X|Y}(X|Y)} > \ell - \tau \right\} + 2^{-\tau}.$$
(c) If we now take $n$ independent pairs $(X_i, Y_i)$ distributed according to $P_{XY}$. What can you say on the limit $\lim_{n \to \infty} \delta_{SW}^{opt}(X^n|Y^n, R^n)$ for different values of $R$?

3. (Advice: Only do this question if you have completed the previous ones) We now consider a different setting called distributed compression. Suppose Alice compresses $X$ using $C_1 : X \to \{0, 1\}^{\ell_1}$ and Charlie compresses $Y$ using $C_2 : Y \to \{0, 1\}^{\ell_2}$ and the decompressor $D : \{0, 1\}^{\ell_1} \times \{0, 1\}^{\ell_2} \to X \times Y$ received both $C_1(X)$ and $C_2(Y)$ and is asked to recover both $X$ and $Y$. In this case the error probability of error if given by $P\{D(C_1(X), C_2(Y)) \neq (X, Y)\}$. We then denote $\delta_{opt}(X, Y, \ell_1, \ell_2)$ to be the smallest error probability that can be achieved. Take $n$ independent pairs $(X_i, Y_i)$ distributed according to $P_{XY}$.

(a) Show that if $R_1 > H(X)$ and $R_2 > H(Y|X)$, then the limit
$$\lim_{n \to \infty} \delta_{opt}(X^n, Y^n, R_1 n, R_2 n) = 0.$$ 

(b) More generally, what can you say on the set $\mathcal{R} \subset \mathbb{R}_+ \times \mathbb{R}_+$ of rates such that
$$\lim_{n \to \infty} \delta_{opt}(X^n, Y^n, R_1 n, R_2 n) = 0$$
for any $(R_1, R_2) \in \mathcal{R}$? (Do not worry about the boundary $\partial \mathcal{R}$ of $\mathcal{R}$). Try to draw schematically the set $\mathcal{R}$. 