Handout 24 Information Theory and Coding
Graded Homework, due Monday December 3 Nov. 20, 2018

You are allowed (even encouraged) to discuss the problems on the homework with your colleagues. However, your solutions should be in your own words. If you collaborated on your solution, write down the name of your collaborators and your sources; no points will be deducted. But similarities in solutions beyond the listed collaborations will be considered as cheating.

PROBLEM 1. Consider a discrete memoryless channel W with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Let

$$\text{enc} : \{1, \ldots, M\} \to \mathcal{X}^n$$

be the encoding function of a block code of blocklength n with M codewords. Let us denote by $x(m, i)$ the i'th letter of the m'th codeword, i.e.,

$$\text{enc}(m) = \big(x(m, 1), \ldots, x(m, n)\big), \qquad m = 1, \ldots, M.$$

We define the *empirical distribution* $p_{\hat{X}}$ of the encoding function as

$$p_{\hat{X}}(x) = (nM)^{-1} \sum_{m=1}^{M} \sum_{i=1}^{n} \mathbb{1}\{x(m, i) = x\}.$$

(a) Verify that $p_{\hat{X}}$ is a probability distribution on $\mathcal{X}$.

Let U be a random variable uniformly distributed on $\{1, \ldots, M\}$. Let $X^n = \text{enc}(U)$. Let Y^n denote the output of the channel W when the channel input is X^n.

(b) Find $p_{X_i}(x)$, the probability $\Pr(X_i = x)$, in terms of $(x(1, i), \ldots, x(M, i))$.

(c) What is the relationship between $p_{\hat{X}}$ and $p_{X_1}, \ldots, p_{X_n}$?

(d) Let $f(\text{enc})$ denote the value of $\frac{1}{n} I(U; Y^n)$, the normalized mutual information between the message U and the channel output Y^n.

Show that $f(\text{enc}) \le I(\hat{X}; Y)$, where $\hat{X}$ is a random variable with distribution $p_{\hat{X}}$ above, and Y is the output of channel W when the channel input is $\hat{X}$.

Suppose now that the encoding function enc is chosen randomly by choosing $(x(m, i) : 1 \le m \le M,\ 1 \le i \le n)$ as i.i.d. random variables with distribution p_X. We denote by Enc the randomly chosen encoder. Note that since the encoder is random, the distribution $p_{\hat{X}}(x)$ is also random.

(e) Show that $E[p_{\hat{X}}(x)] = p_X(x)$.

(f) Show that $E[f(\text{Enc})] \le I(X; Y)$, where f is as in (d), X has distribution p_X, and Y is the output of channel W when the channel input is X.

PROBLEM 2. Consider a discrete memoryless channel with input x and two outputs Y and Z. The channel is described by specifying $W(y, z|x)$ — the probability that $(Y, Z) = (y, z)$ when the channel input is x — for every triple x, y, z.

The output Y is observed at a receiver to whom the transmitter (controlling x) wants to convey a message U. We take U to be random variable uniformly distributed on $\{1, \ldots, M\}$. However the output Z is observed at an eavesdropper from whom the message U should be kept secret.

In this dual task, the encoder can make use of an auxiliary source of randomness Q, which is a random variable uniformly distributed on $\{1, 2, \ldots, J\}$ and is independent of U. The encoder is free to choose the integer J.

In such a scenario, the transmitter is described by an encoding function

$$\text{enc} : \{1, \ldots, M\} \times \{1, \ldots, J\} \to \mathcal{X}^n.$$

The encoding function associates to each possible pair u, q of values of U and Q, a channel input sequence $\text{enc}(u, q) = \big(x(u, q, 1), \ldots, x(u, q, n)\big)$.

The rate of this encoder is (as usual) $\frac{1}{n} \log M$. We measure how much information the eavesdropper has learned about U by the normalized mutual information $\frac{1}{n} I(U; Z^n)$.

The receiver will recover from Y^n an estimate $V = \text{dec}(Y^n)$ of the message. As in class

$$\text{dec} : \mathcal{Y}^n \to \{0, 1, \ldots, M\}$$

is the decoding function with which the estimate is computed. Let $P_e = \Pr(V \neq U)$ denote the average probability of error.

(a) Show that $I(U; Z^n) = I(UQ; Z^n) - I(Q; Z^n U)$.

(b) Show that $\frac{1}{n} I(UQ; Z^n) \leq I(\hat{X}; Z)$. (With $\hat{X}$ as in problem 1 above.)

Fix R and R_0. Set $M = 2^{nR}$ (so that the encoder has rate R) and $J = 2^{nR_0}$. Fix a distribution p_X on the channel input alphabet $\mathcal{X}$.

Choose the encoding function randomly, by choosing $\big(x(m, q, i) : 1 \leq m \leq M,\ 1 \leq q \leq J,\ 1 \leq i \leq n\big)$ i.i.d. with distribution p_X. Let (X, Y, Z) have distribution $p_X(x) W(y, z|x)$.

With this choice quantities such as P_e, $I(U; Z^n)$, $I(Q; Z^n U)$, $I(\hat{X}; Z)$, etc., become random variables.

(c) Suppose $R + R_0 < I(X; Y)$. Show that $E[P_e]$ approaches 0 as n gets large. [Hint: consider a decoder $\text{dec}' : \mathcal{Y}^n \to \{0, 1, \ldots, MJ\}$ that estimates the pair (U, Q) from Y^n. Clearly, P_e is upper bounded by the error probability of this more ambitious decoder.]

(d) Suppose $R_0 < I(X; Z)$. Consider estimating Q from the pair (Z^n, U). Let $T = \text{dec}_0(Z^n, U)$ be this estimate, and $P_0 = \Pr(Q \neq T)$ be the error probability of this estimation. Show that for an appropriate choice of the function dec_0, the expected value $E[P_0]$ of the error probability approaches zero as n gets large.

(e) Again supposing $R_0 < I(X; Z)$, show that $\delta_n = \frac{1}{n} E[H(Q|Z^n U)]$ approaches 0 as n ges large. [Hint: with T and P_0 as in (d), Fano's inequality says $H(Q|T) \leq P_0 \log(J - 1) + h_2(P_0)$.]

(f) With δ_n as in (e), show that $\frac{1}{n} E[I(U; Z^n)] \leq I(X; Z) - R_0 + \delta_n$. [Hint: (a), (b), Problem 1(f).]

(g) Show that for any $R < I(X; Y) - I(X; Z)$ and for any $\epsilon > 0$, there is an enc and dec such that $P_e < \epsilon$ and $\frac{1}{n} I(U; Z^n) < \epsilon$). [Hint: Choose $R_0 = I(X; Z) - \epsilon/4$, and use (c) and (f) to conclude that for large n, $E[P_0] < \epsilon/2$ and $\frac{1}{n} E[I(U; Z^n)] < \epsilon/2$.]