ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 25 | Information Theory and Coding

Graded Homework Solutions | Dec. 4, 2018

PROBLEM 1.

(a) It is clear that $p_{\hat{X}}(x) \geq 0$. Since for each m and i we have $\sum_x \mathbb{1}\{x(m,i) = x\} = 1$, we find that $\sum_x p_{\hat{X}}(x) = 1$, thus verifying that $p_{\hat{X}}$ is a probability distribution on $\mathcal{X}$.

(b) $\Pr(X_i = x) = \sum_{m=1}^{M} \Pr(X_i = x, U = m) = \sum_m \Pr(U = m) \Pr(X_i = x | U = m) = \frac{1}{M} \sum_m \mathbb{1}\{x(i,m) = x\}$.

(c) From (b), $p_{\hat{X}}(x) = \frac{1}{n} \sum_{i=1}^{n} p_{X_i}(x)$, i.e., $p_{\hat{X}}$ is the average of $p_{X_1}, \dots, p_{X_n}$.

(d) By the data processing inequality, $I(U; Y^n) \leq I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n)$. Since the channel is memoryless, $H(Y^n|X^n) = \sum_i H(Y_i|X_i)$. Moreover, $H(Y^n) \leq \sum_i H(Y_i)$. Thus, $\frac{1}{n} I(U; Y^n) \leq \frac{1}{n} I(X^n; Y^n) \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_i)$. Writing $I(X_i, Y_i) = J(p_{X_i}, W)$, we know from class that J is a concave function of its first argument. From (b) $\frac{1}{n} \sum_i p_{X_i} = p_{\hat{X}}$, so, by the concavity of J we have $\frac{1}{n} \sum_i J(p_{X_i}, W) \leq J(p_{\hat{X}}, W) = I(\hat{X}; Y)$.

(e) Observe that $E[\mathbb{1}\{X(m,i) = x\}] = \Pr(X(m,i) = x) = p_X(x)$. It then follows that $E[p_{\hat{X}}(x)] = (nM)^{-1} \sum_m \sum_i p_X(x) = p_X(x)$. (The same argument also shows that for each i, $E[p_{X_i}(x)] = p_X(x)$.)

(f) In (d) we had seen that $f(\text{enc}) \leq J(p_{\hat{X}}, W)$. From the concavity of J in its first argument, it follows that $E[f(\text{Enc})] \leq J(E[p_{\hat{X}}], W) = J(p_X, W) = I(X, Y)$.

The main message of the problem is in (d): to operate at rate R and small probability of error, a code must have a $p_{\hat{X}}$ for which $I(\hat{X}; Y) \geq R$. In particular, a necessary (but not sufficient) condition for reliable communication at rates close to channel capacity is for $p_{\hat{X}}$ to be close to a capacity achieving input distribution.

PROBLEM 2.

(a) By the chain rule $I(UQ; Z^n) = I(U; Z^n) + I(Q; Z^n|U)$. Since $I(Q; U) = 0$, again by the chain rule, $I(Q; Z^n U) = I(Q; Z^n|U)$, so $I(UQ; Z^n) = I(U; Z^n) + I(Q; Z^n U)$.

(b) Note that $(U, Q) \multimap X^n \multimap Z^n$, with X^n determined from (U, Q) by the encoder and Z^n determined from X^n by the channel. Consequently (U, Q), X^n and Z^n play the roles of U, X^n and Y^n in problem 1. We thus obtain from 1(d) that $\frac{1}{n} I(UQ; Z^n) \leq I(\hat{X}; Z)$.

(c) Note that from a decoder dec$'$ that estimates (U, Q) we can obtain a decoder dec that estimates U by throwing away the estimate of Q. Also, as $\Pr(\hat{U} \neq U) \leq \Pr((\hat{U}, \hat{Q}) \neq (U, Q))$, the new decoder dec has a smaller probability of error than dec$'$.

With (U, Q) thought as the 'message', $R + R_0$ is the communication rate (since $\frac{1}{n} \log(MJ) = R + R_0$). From the class we know that as long as the rate is less than $I(X; Y)$, the expected error probability of a randomly chosen code — with each letter of each codeword independently chosen according to distribution p_X — and decoder dec$'$ will approach zero as n gets large. By the remarks in the previous paragraph the same holds for the decoder dec.

(d) As the decoder is provided with the value u of U, it knows that one of J codewords — $\mathrm{enc}(1,u),\ldots,\mathrm{enc}(J,u)$ — is the codeword sent by the transmitter. These J codewords form a code of rate $\frac{1}{n}\log J = R_0$. As these codewords were chosen via the random coding construction, we know from class that as long as $R_0 < I(X;Z)$ the expected error probability $E[P_0]$ (of estimating Q from Z^n and U) appraoches 0 as n gets large.

(e) Since T is a function of (Z^n, U), we have $H(Q|Z^nU) \le H(Q|T) \le P_0\log(J-1) + h_2(P_0)$. As $\log(J-1) \le nR_0$ and $h_2(P_0) \le 1$, we find $\delta_n = \frac{1}{n}E[H(Q|Z^nU)] \le E[P_0]R_0 + \frac{1}{n}$. By (d) $E[P_0]$ approaches zero as n gets large. We conclude that δ_n approaches zero too.

(f) From (a) we know $\frac{1}{n}I(U;Z^n) = \frac{1}{n}I(UQ;Z^n) - \frac{1}{n}I(Q;Z^nU)$. From (b) and 1(f), we have $\frac{1}{n}E[I(UQ;Z^n)] \le I(X;Z)$. From (e), we have $\frac{1}{n}E[I(Q;Z^nU)] = \frac{1}{n}E[H(Q) - H(Q|Z^nU)] = R_0 - \delta_n$. Putting these together, we find $\frac{1}{n}E[I(U;Z^n)] \le I(X;Z) - R_0 + \delta_n$.

(g) Since $R < I(X;Y) - I(X;Z)$, choosing $R_0 = I(X;Z) - \epsilon/4$ will ensure that $R + R_0 < I(X;Y)$ as well as $R_0 < I(X;Z)$. Thus from (e) and (b), by choosing n large enough we can ensure $\delta_n < \epsilon/4$ and $E[P_e] < \epsilon/2$. We thus obtain from (f) that $E[P_e + \frac{1}{n}I(U;Z^n)] < \epsilon$. Consequently, there must exist an (enc,dec) pair such that $P_e + \frac{1}{n}I(U;Z^n) < \epsilon$, which implies that both P_e and $\frac{1}{n}I(U;Z^n)$ are smaller than ϵ.

The setup we examined in this problem is known as the Wiretap Channel, where an eavesdropper observing Z has to be kept ignorant of the message U while reliably communicating the message to the legitimate receiver who observes Y. It is possible to show a stronger result than we proved here: when $R < I(X;Y) - I(X;Z)$ we can make $I(U;Z^n)$ close to zero (without the normalization by n).

Under further assumptions (e.g., $X \multimap Y \multimap Z$), it is possible to show a converse: if $R > \max_{p_X}[I(X;Y) - I(X;Z)]$, then $\frac{1}{n}I(U;Z^n)$ cannot be made arbitrarily small.