ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 31 Information Theory and Coding
Solutions to Final Exam Jan. 19, 2016

PROBLEM 1.

(a) By the chain rule we have

$$\begin{aligned} H(X^n, Z^n) &= H(X^n) + H(Z^n|X^n) \\ &= H(Z^n) + H(X_1|Z^n) + H(X_2^n|X_1, Z^n). \end{aligned}$$

(where $X_2^n = (X_2, \dots, X_n)$). Since (X_1, Z^n) determines X_2^n (because, X_1 determines $\hat{X}_2$ and, knowing $(\hat{X}_2, Z_2)$, X_2 is known, then (X_1, X_2) determines $\hat{X}_3$ which, together with Z_3 determines X_3, ...), $H(X_2^n|X_1, Z^n) = 0$. Therefore,

$$H(Z^n) = H(X^n) - H(X_1|Z^n)$$

which (together with the fact that $H(X_1|Z^n) \le H(X_1) \le 1$) implies

$$H(X^n) - 1 \le H(Z^n) \le H(X^n)$$

Consequently

$$\mathcal{H}(Z) = \lim_{n\to\infty} \frac{1}{n} H(Z^n) = \lim_{n\to\infty} \frac{1}{n} H(X^n) = \mathcal{H}(X).$$

(b) Once again using the chain rule,

$$\begin{aligned} H(Z_n, X_n|X^{n-1}) &= H(Z_n|X^{n-1}) + H(X_n|Z_n, X^{n-1}) \\ &= H(X_n|X^{n-1}) + H(Z_n|X^n), \end{aligned}$$

and since $H(X_n|Z_n, X^{n-1}) = H(Z_n|X^n) = 0$ (for the reasons explained in (a)),

$$h_2(p_n) = H(Z_n) \ge H(Z_n|X^{n-1}) = H(X_n|X^{n-1}).$$

(c) From (b) and since X is a stationary process we have

$$\liminf_{n\to\infty} h_2(p_n) \ge \lim_{n\to\infty} H(X_n|X^{n-1}) = \mathcal{H}(X).$$

Moreover, there exists a subsequence of $\{p_n : n = 1, 2, \dots\}$, say $\{p_{n_i} : n_i \in \mathbb{N}, i = 1, 2, \dots\}$ for which $\liminf_{n\to\infty} p_n = \lim_{i\to\infty} p_{n_i}$. Thus, by the continuity of $h_2(\cdot)$,

$$\liminf_{n\to\infty} h_2(p_n) = \lim_{i\to\infty} h_2(p_{n_i}) = h_2(\liminf_{i\to\infty} p_{n_i}) = h_2(p).$$

PROBLEM 2.

(a) Since the channel is memoryless and feedback-free transmission is assumed, from the code construction, it is obvious that $(\mathrm{enc}_1(m_1), \mathrm{enc}_2(m_2), Y^n)$ is an i.i.d. length-n sequence of (X_1, X_2, Y)'s drawn from distribution $p(x_1, x_2, y) = p_1(x_1)p_2(x_2)p(y|x_1, x_2)$. Therefore, for sufficiently large n, the probability of this sequence being ϵ-typical is as high as desired.

(b) Now, $(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n)$ is an i.i.d. sequence (of length n) whose components are distributed according to $p_1(x_1)p(y, x_2)$ where $p(y, x_2) = \sum_{x_1'} p_1(x_1')p_2(x_2)p(y|x_1', x_2)$.

(c) $\Pr\{(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n) \in T\}$ is the probability of a length n i.i.d. sequence X_1^n whose elements have distribution p_1 being jointly ϵ-typical (with respect to the distribution $p_1(x_1)p(y, x_2|x_1)$ where $p(y, x_2|x_1) = p(x_2)p(y|x_1, x_2)$) with an independent length n sequence of (X_2, Y)'s drawn from distribution $p(y, x_2)$ (defined in (b)). Thus,

$$\Pr\{(\text{enc}_1(\tilde{m}_1), \text{enc}_2(m_2), Y^n) \in T\} \doteq 2^{-nI(X_1, X_2 Y)}.$$

(In the course we have seen this result for two random variables X and Y; it is obvious that we can replace X by X_1 and Y by (X_2, Y) to derive the above result).

(d) From (a) we know that the probability of the correct message m_1 not being in the list of typical m_1's at decoder 2 is small, say at most $\epsilon/2$.

From (c), the probability of each incorrect $\tilde{m}_1$ being on that list (at decoder 2) is equal (up to sub-exponential factors) to $2^{-nI(X_1;X_2Y)}$. Since there are $M - 1 \leq 2^{nR_1}$ such $\tilde{m}_1$'s, the probability of having *an* incorrect message on the list is, by the union bound, at most $2^{n[R_1 - I(X_1;X_2Y)]}$ which is exponentially small in n provided that $R_1 < I(X_1; X_2Y)$. Thus, for large enough n, this probability is also smaller than $\epsilon/2$.

Consequently, the average probability of decoding error at decoder 2 is at most ϵ provided that $R_1 < I(X_1, X_2Y)$.

By symmetry, the average probability of decoding error at decoder 1 is smaller than ϵ if $R_2 < I(X_2, X_1, Y)$.

Since the average probability of error (over the generation of codebooks) is small (for rate pairs (R_1, R_2) satisfying $R_1 < I(X_1; Y, X_2)$ and $R_2 < I(X_2; Y, X_1)$), there exists a pair of codebooks of rates (R_1, R_2) in the ensemble for which the average error probability is small, thus such (R_1, R_2)'s are achievable.

(e) Firstly note that since X_1 and X_2 are independent, $I(X_1; Y X_2) = I(X_1; Y|X_2)$ (similarly $I(X_2; Y X_1) = I(X_2; Y|X_1)$).

Since $Y = X_1 \times X_2$, conditioned on $\{X_2 = 0\}$, Y contains no information about X_1, whereas conditioned on $\{X_2 = 1\}$, $Y = X_1$. Thus, assuming $\Pr\{X_1 = 1\} = p_1$ and $\Pr\{X_2 = 1\} = p_2$,

$$\begin{aligned} I(X_1; Y|X_2) &= \Pr\{X_2 = 0\}I(X_1; Y|X_2 = 0) + \Pr\{X_2 = 1\}I(X_1; Y|X_2 = 1) \\ &= 0 + p_2 h_2(p_1) \end{aligned}$$

where $h_2(\cdot)$ is the binary entropy function. Similarly it follows that

$$I(X_2; Y|X_1) = p_1 h_2(p_2).$$

Suppose $p_1 = p_2 = p$, then all rates (R_1, R_2) satisfying

$$R_1 < p h_2(p) \qquad R_2 < p h_2(p)$$

are achievable. In particular, $p h_2(p) \geq \frac{1}{2}$ for some $p \geq \frac{1}{2}$ (it evaluates to $\frac{1}{2}$ at $p = \frac{1}{2}$ but it is increasing, so it will go above $\frac{1}{2}$ as p increases). The set of achievable rate pairs corresponding to such p's violate $R_1 + R_2 < 1$.

Problem 3.

(a) For large enough n, (X^n, Y^n) is jointly ϵ-typical with respect to the distribution $p(x, y)$. Thus, the chance of the true sequence X^n not appearing on Bob's list vanishes as n gets large.

(b) For an ϵ-typical sequence y^n, $p(y^n) \leq 2^{-n(1-\epsilon)H(Y)}$. Similarly, for jointly ϵ-typical sequences (x^n, y^n), $p(x^n, y^n) \geq 2^{-n(1+\epsilon)H(X,Y)}$. We, then, have

$$
\begin{aligned}
2^{-n(1-\epsilon)H(Y)} &\geq p(y^n) \\
&= \sum_{x^n} p(x^n, y^n) \\
&\geq \sum_{\substack{x^n: \\ (x^n, y^n) \in T_\epsilon}} p(x^n, y^n) \\
&\geq |\{x^n : (x^n, y^n) \in T_\epsilon\}| 2^{-n(1+\epsilon)H(X,Y)},
\end{aligned}
$$

where T_ϵ denotes the set of jointly typical (x^n, y^n)'s. Therefore, by noticing that $(1+\epsilon)H(X,Y) - (1-\epsilon)H(Y) = (1+\epsilon)H(X|Y) + 2\epsilon H(Y)$, for an ϵ-typical y^n,

$$|\{x^n : (x^n, y^n) \in T_\epsilon\}| \leq 2^{n[(1+\epsilon)H(X|Y) + 2\epsilon H(Y)]} \approx 2^{nH(X|Y)}.$$

(c) Given a typical sequence y^n, x^n appears on Bob's list if (x^n, y^n) is typical and $\text{label}(x^n) = \text{label}(x_0^n)$ (where x_0^n is the true sequence which we assume to be typical as well – otherwise Bob will not receive any label from Alice). The number of wrong sequences is thus

$$N_w(x_0^n, y^n) := \sum_{\substack{x^n: \\ (x^n, y^n) \in T_\epsilon \\ x^n \neq x_0^n}} \mathbf{1}\{\text{label}(x^n) = \text{label}(x_0^n)\}.$$

Since the labels are assigned independently and uniformly from $\{1, \ldots, 2^{nR}\}$,

$$E[\mathbf{1}\{\text{label}(x^n) = \text{label}(x_0^n)\}] = \Pr\{\text{label}(x^n) = \text{label}(x_0^n)\} = 2^{-nR} \qquad (\text{if } x^n \neq x_0^n).$$

Consequently, using (b) we have:

$$E[N_w(x_0^n, y^n)] = |\{x^n : (x^n, y^n) \in T_\epsilon\} \setminus \{x_0^n\}|\, 2^{-nR} \leq 2^{-n[R - (H(X|Y) + \delta)]},$$

for some $\delta = \delta(\epsilon)$ which goes to 0 as $\epsilon \to 0$.

(d) A decoding error will happen if either (X^n, Y^n) are atypical or they are typical but Bob's list has more than one element. In other words,

$$\Pr\{\text{error}\} = \Pr\{(X^n, Y^n) \notin T_\epsilon\} + \Pr\{(X^n, Y^n) \in T_\epsilon, N_w(X^n, Y^n) \geq 1\}.$$

The first term on the right-hand-side of the above goes to 0 as n gets large (independent of R). For the second term we have

$$
\begin{aligned}
&\Pr\{(X^n, Y^n) \in T_\epsilon, N_w(X^n, Y^n) \geq 1\} \\
&\quad = \Pr\{(X^n, Y^n) \in T_\epsilon\} \Pr\{N_w(X^n, Y^n) \geq 1 | (X^n, Y^n) \in T_\epsilon\} \\
&\quad \leq \Pr\{N_w(X^n, Y^n) \geq 1 | (X^n, Y^n) \in T_\epsilon\} \\
&\quad \overset{(*)}{\leq} E[N_w(X^n, Y^n) | (X^n, Y^n) \in T_\epsilon] \\
&\quad \leq 2^{-n[R - H(X|Y) - \delta]},
\end{aligned}
$$

where $(*)$ follows from the Markov inequality.

Thus, if $R > H(X|Y)$ the second term also vanishes as n gets large which means Bob will decide correctly with high probability.

PROBLEM 4.

(a) Let $\mathbf{x}'$ and $\mathbf{y}'$ be two codewords in $\mathcal{C}'$ corresponding to information vectors $\mathbf{u} = (u_0, u_1, \ldots, u_{k-1})$ and $\mathbf{v} = (v_0, \ldots, v_{k-1})$ respectively. $\alpha\mathbf{x}' + \beta\mathbf{y}'$, $\alpha \in \mathbb{F}, \beta \in \mathbb{F}$ corresponds to the encoding of $\alpha\mathbf{u} + \beta\mathbf{v} = (\alpha u_0 + \beta v_0, \alpha u_1 + \beta v_1, \ldots, \alpha u_{k-1} + \beta v_{k-1})$, hence is a codeword of $\mathcal{C}'$ as well.

(b) The number of zeros in $(x_1, \ldots, x_n)$ is the number of roots the polynomial $u(D) = u_0 + u_1 D + \cdots + u_{k-2} D^{k-2}$ (note that $u_{k-1} = 0$) has among $\{\alpha_1, \ldots, \alpha_n\}$. A polynomial of degree at most $k-2$ has at most $k-2$ roots, and thus a weight of $\text{weight}(x_1, \ldots, x_n) \geq n-(k-2) = n+2-k$. Since $u_{k-1} = 0$, $\text{weight}(u_{k-1}, x_1, \ldots, x_n) = \text{weight}(x_1, \ldots, x_n) \geq n + 2 - k$.

(c) Since $u_{k-1} \neq 0$, $\text{weight}(u_{k-1}, x_1, \ldots, x_n) = 1 + \text{weight}(x_1, \ldots, x_n)$. Now among $x_1, \ldots, x_n$ at most $k-1$ elements can be zero (since they are evaluations of a polynomial of degree $k-1$), hence $\text{weight}(x_1, \ldots, x_n) \geq n+1-k$. Thus, $\text{weight}(\mathbf{x}') \geq n+2-k$.

(d) From (a), (b) and (c) we have $d_{\min}(\mathcal{C}') = \min_{\mathbf{x}' \in \mathcal{C}'} \text{weight}(\mathbf{x}') \geq n + 2 - k$. On the other hand, the Singleton bound states that for any linear code of blocklength $n + 1$ and dimension k, $d_{\min} \leq n - k + 2$. This shows the code $\mathcal{C}'$ has minimum distance exactly equal to $n - k + 2$.