ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 32 Information Theory and Coding

Final Exam Solutions Jan. 18, 2014

Problem 1.

(a) Given the dictionaries $\mathcal{D}_1$, $\mathcal{D}_2$ and the decision rule in (ii), we are sure that W will belong to a set of words $\mathcal{D}$, and b needs to be $\lceil \log |\mathcal{D}| \rceil$. We now need to determine how small and how large $\mathcal{D}$ may be. E.g., if the rule is "longer" then $\mathcal{D}$ is obtained from $\mathcal{D}_1 \cup \mathcal{D}_2$ by removing words that are prefix of another; if the rule is "shorter" $\mathcal{D}$ is obtained from $\mathcal{D}_1 \cup \mathcal{D}_2$ by removing words which have another word as a prefix; if rule is "coin flip" $\mathcal{D}$ is equal to $\mathcal{D}_1 \cup \mathcal{D}_2$. Clearly, $\mathcal{D}$ is a subset $\mathcal{D}_1 \cup \mathcal{D}_2$, and may be equal in the case of the "coin flip" rule. Thus $|\mathcal{D}| \le |\mathcal{D}_1| + |\mathcal{D}_2| = 2M$, (and equality is possible when $\mathcal{D}_1$ and $\mathcal{D}_2$ are disjoint, e.g., $\mathcal{D}_1 = \{\mathtt{a}, \mathtt{ba}, \mathtt{bb}\}$, $\mathcal{D}_2 = \{\mathtt{aa}, \mathtt{ab}, \mathtt{b}\}$.)

On the other hand since $\mathcal{D}_1$ and $\mathcal{D}_2$ are valid, so is $\mathcal{D}$: every infinite sequence $U_1U_2\dots$ has a prefix in $\mathcal{D}$. Consequently $|\mathcal{D}|$ is at least $|\mathcal{U}|$, and equality is possible by the example given in the hint. Thus

$$|\mathcal{U}| \le |\mathcal{D}| \le 2M$$

and we find $\lceil \log |\mathcal{U}| \rceil \le b \le 1 + \lceil \log M \rceil$.

(b) We need to provide binary representations for $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, so, $b = \log |\mathcal{D}_1 \cup \mathcal{D}_2|$. As $M \le |\mathcal{D}_1 \cup \mathcal{D}_2| \le 2M$, we find $\lceil \log M \rceil \le b \le 1 + \lceil \log M \rceil$.

(c) Since we will use $b = \log |\mathcal{D}_1 \cup \mathcal{D}_2|$ bits to describe the word no matter what decision rule is used, to minimize bits per source letter, we need to make sure we describe as many source letters as we can; thus we should use the "longer" rule.

(d) Using the "longer rule", we have $\text{length}(W) \ge \text{length}(W_1)$. Also, from (b) we have $b < 2 + \log M$. Thus

$$\frac{b}{E_1[\text{length}(W)]} < \frac{2 + \log M}{E_1[\text{length}(W_1)]} = \frac{2}{E_1[\text{length}(W_1)]} + \frac{\log M}{E_1[\text{length}(W_1)]}.$$

Since $E_1[\text{length}(W_1)] \ge (\log M)/H_1$, the first term approaches zero as M gets large. The second term approaches H_1 as M gets large, thus

$$\frac{b}{E_1[\text{length}(W)]} \le H_1 + \epsilon$$

where ϵ can be made arbitrarily small by taking M large. The case with $E_2[\cdot]$ and H_2 follow in analogous fashion.

Moral of the story: if two Tunstall codes are designed under different assumptions, they can be combined into a single design that is guaranteed to perform well under either of the assumptions.

Problem 2.

(a) We have $I(W_1W_2;Z^n) \leq I(X^n;Z^n) \leq \sum_i I(X_i;Z_i)$, the first by data processing, the second by the memoryless property of the channel. Since each $I(X_i;Z_i)$ is less than the capacity of BEC(q) and since the capacity of the BEC(q) is $1-q$, we find $I(W_1W_2;Z^n) \leq n(1-q)$.

(b) By the chain rule $I(W_1W_2;Z^n) = I(W_2;Z^n) + I(W_1;Z^n|W_2) \geq I(W_1;Z^n|W_2)$. Also by the chain rule $I(W_1;Z^nW_2) = I(W_1;W_2) + I(W_1;Z^n|W_2) = I(W_1;Z^n|W_2)$. Thus $I(W_1;Z^nW_2) \leq I(W_1W_2;Z^n) \leq n(1-q)$.

(c) By the chain rule $I(W_1W_2;Z^n) = I(W_1;Z^n) + I(W_2;Z^n|W_1)$. But $I(W_2;Z^nW_1) = I(W_2;W_1) + I(W_2;Z^n|W_1) = I(W_2;Z^n|W_1)$. Thus $I(W_1W_2;Z^n) = I(W_1;Z^n) + I(W_2;Z^nW_1)$.

(d) The procedure is identical to the construction of a random code of rate $\frac{1}{n}\log(M_1M_2) = R_1 + R_2$, according to the distribution $p_X(0) = p_X(1) = 1/2$. Note that for this distribution $I(X;Y) = 1-p$.

In class we saw that the expected probability of error (i.e. the probability of wrongly determining (W_1W_2) from Y^n) of a random code constructed as above approaches zero with increasing n if $R_1 + R_2 < I(X;Y) = 1-p$.

(e) In determining W_2 from (Z^n, W_1) we already know the value of W_1, say $W_1 = w_1$. So, all we need is to determine is which of the M_2 codewords $X^n(w_1,1),\dots,X^n(w_1,M_2)$ is transmitted from the observation Z^n. Since these M_2 codewords form a random code of rate R_2 constructed according to the distribution $p_X(0) = p_X(1) = 1/2$, and since $I(X;Z) = 1-q$, the expected probability of error in the determination of W_2 from (Z^n, W_1) approaches zero with increasing n if $R_2 < I(X;Z) = 1-q$.

(f) The conditions $R_1 < q-p$, $R_2 < 1-q$ ensure that $R_1 + R_2 < 1-p$; so from (d) and (e) we conclude that $E[\Delta_n]$ and $E[\Theta_n]$ can both be made arbitrarily small, say, both less than $\epsilon/2$. Thus for a randomly constructed code $E[\Delta_n + \Theta_n] < \epsilon$ for sufficiently large n, and thus there is a code with $\Delta_n + \Theta_n < \epsilon$. For this code both Δ_n and Θ_n are smaller than ϵ.

(g) By Fano's inequality $H(W_2|Z^nW_1)$ is upper bounded by $h_2(p_e) + p_e\log(M_2-1)$ where p_e is the probability of error in determining W_2 from (Z^n, W_1). But $\log(M_2-1) < nR_2$, and for the code in (f) we know that $p_e < \epsilon$. Thus $H(W_2|Z^nW_1) \leq nR_2\epsilon + h_2(\epsilon)$.

From (c) $I(W_1;Z^n) = I(W_1W_2;Z^n) - I(W_2;Z^nW_1) = I(W_1W_2;Z^n) + H(W_2|Z^nW_1) - H(W_2)$. From (a) the first term on the right hand side $I(W_1W_2;Z^n) \leq n(1-q)$; we just upper bounded the second term, and we know $H(W_2) = nR_2$. Thus

$$I(W_1;Z^n) \leq n(1-q-R_2) + nR_2\epsilon + h_2(\epsilon).$$

(h) Given $R_1 < p-q$, and choosing $R_2 = 1-q-\epsilon$ we get from (f) and (g) that there is an encoder for which

$$\Delta_n < \epsilon, \quad \frac{1}{n}I(W_1;Z^n) < R_2\epsilon + \epsilon + \frac{1}{n}h_2(\epsilon) < 3\epsilon$$

for large enough n. As $\epsilon > 0$ is arbitrary this is equivalent to what was to be shown.

Moral of the story: we can design codes up to rate $q-p$ which let W_1 to be sent reliably to the receiver observing Y^n but keep it secret from a wiretapper observing Z^n.

Problem 3.

(a) Given a code $\mathcal{C}$ with M codewords and blocklength n, and $0 \le k \le n$, partition the codewords into 2^k groups according to their first k bits. The group with the largest number of codewords will contain at least $M' = \lceil M/2^k \rceil$ codewords. The minimum distance within that group is upper bounded by $d_0(M', n-k)$ since all codewords in the group agree in their first k bits. Thus the minimimum distance of the code $\mathcal{C}$ is upper bounded by $d_0(\lceil M/2^k \rceil, n-k)$. Since this is true for each $k \in \{0, \dots, n\}$ we conclude that $d_{\min} \le d_1(M, n)$.

(b) With $d_0(M,n) = \begin{cases} n & M \ge 2 \\ \infty & M \le 1 \end{cases}$ the minimum over k is obtained by choosing k as large as possible while keeping $M/2^k > 1$. Thus the bound d_1 says "$d_{\min} \le n-k$ when $M > 2^k$" which is the Singleton bound we derived in class.

(c) Each pair (m, m') contributes 1 to the sum when $a_m = 0$ and $a_{m'} = 1$ or when $a_m = 1$ and $a_{m'} = 0$. There are $M_0 M_1$ pairs of the first type and $M_1 M_0$ pairs of the second type. Thus the sum equals $2M_0M_1$. As $M_0 + M_1 = M$, we have $M_0 M_1 \le M^2/4$, from which the final inequality follows.

(d) As $d_H(\mathbf{x}_m, \mathbf{x}_{m'}) \ge d_{\min}$ for every $m \ne m'$, the first inequality follows by summing both sides. For the second write $d_H(\mathbf{x}_m, \mathbf{x}_{m'}) = \sum_{i=1}^n d_H(x_{mi}, x_{m'i})$ to obtain

$$\sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \ne m}}^{M} d_H(\mathbf{x}_m, \mathbf{x}_{m'}) = \sum_{i=1}^{n} \sum_{m=1}^{M} \sum_{\substack{m'=1 \\ m' \ne m}}^{M} d_H(x_{mi}, x_{m'i}).$$

By (c) for each i the inner double-sum is upper bounded by $M^2/2$ and the conclusion follows.

(e) Applying (d) gives $d_{\min} \le 7$. Chosing k to be either of 6, 7, or 8 improves the bound to $d_{\min} \le 4$.

Moral of the story: any general upper bound on minimum distance should be checked to see if it can be improved by the simple procedure in (a). The bound in (d) (together with its improvement via (a) as done in (e)) is known as the Plotkin bound.