ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 31 Information Theory and Coding

Solutions to Final Exam Jan. 22, 2015

PROBLEM 1.

(a)
$$\begin{aligned} p_{i+1} &= \Pr(U_{i+1} = 0) = \Pr(U_{i+1} = 0, U_i = 0) + \Pr(U_{i+1} = 0, U_i = 1) \\ &= \Pr(U_{i+1} = 0|U_i = 0)\Pr(U_i = 0) + \Pr(U_{i+1} = 0|U_i = 1)\Pr(U_i = 1) \\ &= \Pr(U_{i+1} = 0|U_i = 0)\Pr(U_i = 0) + \Pr(U_{i+1} = 0|U_i = 1)\Pr(U_i = 1) \\ &= (1 - a_i)p_i + b_i(1 - p_i). \end{aligned}$$

(b) By stationarity p_i does not change with i. Also by stationarity $P(U_{i+1} = 1, U_i = 0) = a_i p_i$ does not change with i, thus a_i does not change with i. Similar reasoning holds for b_i.

(c) For stationary processes the entropy rate is given by $\lim_i H(U_i|U^{i-1})$, and we also know that the sequence in the limit is monotone non-increasing. In particular, $H(U_2|U_1)$ is an upper bound on the entropy rate. Furthermore $H(U_2|U_1 = 0) = h_2(a)$, $H(U_2|U_1 = 1) = h_2(b)$, and thus $H(U_2|U_1) = ph_2(a) + (1 - p)h_2(b)$.

(d) For a Markov process, the entropy rate equals $H(U_2|U_1)$, so the upper bound in (c) is the exact value. Thus, among all processes with the same transition probabilities the Markov process has the largest entropy rate.

(e) For such a process we see that $b = 1$, and from part (a) we find that $p = 1/(1+a)$. By (c) and (d) we find that the maximal entropy rate for a given value of the parameter a is $h_2(a)/(1 + a)$. It only remains to maximize this quantity to over the choice of a to find the maximal entropy rate. (Using standard tools of calculus it is easy to show that the maximum is achieved when $a = (3 - \sqrt{5})/2$.)

PROBLEM 2.

(a) The difference between the left and right sides is

$$\sum_{x,y} Q^*(x)W(y|x)\log\frac{P^*(y)}{P(y)} = \sum_y P^*(y)\log\frac{P^*(y)}{P(y)} = D(P^*\|P) \geq 0.$$

(b) The left hand side of (a), is upper bounded by $\max_x \sum_y W(y|x)\log\frac{W(y|x)}{P(y)}$ whereas the right hand side of (a) equals C.

(c) The Kuhn-Tucker conditions for a capacity achieving input distribution Q^* were derived in class to be

$$\sum_y W(y|x)\log\frac{W(y|x)}{P^*(y)} \leq C, \quad \text{for all } x$$

with equality whenever $Q^*(x) > 0$. Consequently, $\max_x \sum_y W(y|x)\log\frac{W(y|x)}{P^*(y)} = C$.

(d) With $f(Q) = \max_x \sum_y W(y|x)\log\frac{W(y|x)}{P(y)}$, from (b) we see that $f(Q) \geq C$ and that $f(Q^*) = C$. Thus $C = \min_Q f(Q)$.

Problem 3.

(a) Note that $Y = 1$ if and only if $X = 1$ and the channel does not flip. Thus, $\Pr(Y = 1) = (1-p)/2$. An incompatibility between $\tilde{X}$ and Y occurs if only if $\tilde{X} = 0$ and $Y = 1$. Since these two events are independent $\alpha(p) = 1 - (1-p)/4 = (3+p)/4$. Furthermore, since $\tilde{X}$ and Y are indpendent, conditining on Y does not change the distribution of $\tilde{X}$; Thus $\beta(p) = 1/2$.

(b) Since $(\tilde{X}_i, Y_i)$ are i.i.d., and since for each i $(\tilde{X}_i, Y_i)$ is compatible with probability $\alpha(p)$, we see that $\tilde{X}^n$ and Y^n will be compatible with probability $\alpha(p)^n$.

(c) Without loss of generality assume that $Y_1 = \cdots = Y_k = 1$ and the remaining Y_i's are 0. Since when $Y_i = 0$ any value of $\tilde{X}_i$ is compatible, we see that $\tilde{X}^n$ is compatible with $\tilde{Y}^n$ if and only if $\tilde{X}_1 = \ldots = \tilde{X}_k = 1$. By the independence of $\tilde{X}^n$ from Y^n, this event has probability $\beta(p)^k = 2^{-k}$.

(d) Since for the correct message m, $X^n(m)$ is always compatible with Y^n, the receiver will make an error if and only if one of the $M-1$ incorrect messages is compatible with Y^n. By (b) for each of these incorrect messages the probability of being compatible with Y^n is $\alpha(p)^n$, and by the union bound the error probability is upper bounded by

$$(M-1)\alpha(p)^n < 2^{nR}\alpha(p)^n = 2^{n(R+\log\alpha(p))}$$

which approaches zero as long as $R < R_0 = -\log\alpha(p)$.

(e) Let us compute the error probability conditional on the number of 1's, K, in Y^n. By (c), conditional on $K = k$, each of incorrect codewords has a probability $\beta(p)^k$ of being compatible with Y^n, so, using the union bound, the probability of error, conditional on k 1's in Y^n is upper bounded by

$$(M-1)\beta(p)^k < 2^{nR}\beta(p)^k$$

Also note that Y_i are i.i.d., with $\Pr(Y_i = 1) = (1-p)/2$. Consequently, by the law of large numbers for any $q < (1-p)/2$, we have $\Pr(K < nq) \to 0$. We can now write

$$\begin{aligned}\Pr(\text{Error}) &= \Pr(\text{Error}|K < nq)\Pr(K < nq) + \Pr(\text{Error}|K \geq nq)\Pr(K \geq nq)\\ &\leq \Pr(K < nq) + \Pr(\text{Error}|K \geq nq).\end{aligned}$$

The first term decays to zero with increasing n as long as $q < (1-p)/2$, and by the computation before, the second term, $\Pr(\text{Error}|K > nq)$ is upper bounded by $2^{n(R+q\log\beta(p))}$ which decays to zero as long as $R < -q\log\beta(p)$. Consequently whenever $R < R_1 = -\frac{1-p}{2}\log\beta(p) = \frac{1-p}{2}\log 2$, the error probability will approach zero with increasing n.

Problem 4.

(a) If $Z_i = M$ there is nothing to prove. Otherwise there is a codework $\mathbf{x}'$ for which $\mathbf{x}'_i = 1$. Note now that for any codeword $\mathbf{x}$, by the linearity of $\mathcal{C}$, $\mathbf{x}' + \mathbf{x}$ is also a codeword, and thus the map $\mathbf{x} \to \mathbf{x} + \mathbf{x}'$ is a bijection from $\mathcal{C}$ to $\mathcal{C}$. Furthermore because $\mathbf{x}'_i = 1$, this bijection flips the i'th component of $\mathbf{x}$. Consequently there are as many codewords with $\mathbf{x}_i = 0$ as with $\mathbf{x}_i = 1$, and so $Z_i = M/2$.

(b) Note that $I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n)$. By the channel being memoryless $H(Y^n|X^n) = \sum_i H(Y_i|X_i)$. On the other hand, $H(Y^n) \leq \sum_i H(Y_i)$ with equality if and only if $\{Y_i\}$ are independent. Thus,

$$I(X^n; Y^n) \leq \sum_i H(Y_i) - H(Y_i|X_i) = \sum_i I(X_i; Y_i).$$

(c) With X^n chosen uniformly from $\mathcal{C}$, by (a) we see that for each i either $\Pr(X_i = 0) = 1$ (in which case $I(X_i; Y_i) = 0$) or $\Pr(X_i = 0) = 1/2$, (in which case $I(X_i; Y_i) = I(W)$.

(d) By (b) and (c) we see that $I(X^n; Y^n) \leq nI(W)$. Suppose now reliable communication were possible at a rate R using linear codes. Thus for any $\epsilon > 0$, there is a linear code with error probability at most $\epsilon > 0$, and rate at least R. By Fano's inequality, the mutual information between the input message and the decoded message is at least $nR(1 - \epsilon) - h_2(\epsilon)$. By the data processing theorem

$$nR(1 - \epsilon) - h_2(\epsilon) \leq I(X^n; Y^n) \leq nI(W),$$

and thus $R \leq I(W) + \epsilon + \frac{1}{n} h_2(\epsilon)$. Since this is true for every $\epsilon > 0$ and since $h_2(\epsilon) \to 0$ as $\epsilon \to 0$ we see that $R \leq I(W)$.