ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 29 Information Theory and Coding

Solutions to Final Exam Jan. 30, 2012

PROBLEM 1.

(a) Since the events are independent $\Pr(\cap_i E_i) = \prod_i \Pr(E_i) = (1-p)^M$. From the hint $1-p \le e^{-p}$; thus $(1-p)^M \le \exp(-pM)$.

(b) The decoder output is not equal to encoder input if all the events $E_i = \{\mathbf{U}(i) \neq \mathbf{u}\}$ happen. Since $\mathbf{U}(i)$ are chosen independently these events are independent and each has probability $1 - p_{U^n}(\mathbf{u})$. The claim now follows from part (a) with $p = p_{U^n}(\mathbf{u})$.

(c) If $\mathbf{u}$ is ϵ-typical

$$p_{U^n}(\mathbf{u}) = \prod_a p_U(a)^{\#\{i:u_i=a\}} \ge \prod_a p_U(a)^{n(1+\epsilon)p_U(a)} = 2^{-n(1+\epsilon)H(U)}.$$

The claim now follows from part (c).

(d) Given $R > H(U)$ we can pick $\epsilon > 0$ such that $(1+\epsilon)H(U) < R$. Let T_ϵ^n denote the ϵ-typical sequences. The probability of error can be written as

$$\begin{aligned}\Pr(\operatorname{Dec}\operatorname{Enc}(\mathbf{U})) \neq \mathbf{U}) = \Pr(\operatorname{Dec}\operatorname{Enc}(\mathbf{U})) \neq \mathbf{U} | \mathbf{U} \in T_\epsilon^n) \Pr(\mathbf{U} \in T_\epsilon^n) \\ + \Pr(\operatorname{Dec}\operatorname{Enc}(\mathbf{U})) \neq \mathbf{U} | \mathbf{U} \notin T_\epsilon^n) \Pr(\mathbf{U} \notin T_\epsilon^n)\end{aligned}$$

The first term is less than $\Pr(\operatorname{Dec}\operatorname{Enc}(\mathbf{U}) \neq \mathbf{U} | \mathbf{U} \in T_\epsilon^n)$, which, in turn, is less than $\exp(-2^{n[R-(1+\epsilon)H(U)]})$ by part (c). The second term is less than $\Pr(\mathbf{U} \notin T_\epsilon^n)$. Both of these go to zero as n gets large.

PROBLEM 2.

(a) Taking the hint, if $0 < \beta_1 < \beta_2$ we can write $\beta_1 = (1-\lambda)0 + \lambda\beta_1$ with $\lambda = \beta_1/\beta_2$. Since $\lambda \in (0,1)$ and since f is concave

$$f(\beta_1) \ge (1-\lambda)f(0) + \lambda f(\beta_2) = \beta_1 f(\beta_2)\beta_1$$

thus $f(\beta_1)/\beta_1 \ge f(\beta_2)/\beta_2$.

(b) Recall that $C(\beta) = \max_{p_X : E[X] \le \beta} I(X;Y)$. Since $E[X] = \Pr(X = 1)$, we see that $C(\beta) = \max_{\alpha \le \beta} I(\alpha)$.

Furthermore, since $I(\cdot)$ is concave, by denoting by β_0 the value of β that maximizes $I(\beta)$, we see that $I(\cdot)$ is increasing on $[0, \beta_0]$ and decreasing on $[\beta_0, 1]$. Consequently, we can write

$$C(\beta) = \begin{cases} I(\beta) & \beta \le \beta_0 \\ I(\beta_0) & \beta \ge \beta_0. \end{cases}$$

(c) From part (b)

$$C(\beta)/\beta = \begin{cases} I(\beta)/\beta & \beta \le \beta_0 \\ I(\beta_0)/\beta & \beta \ge \beta_0. \end{cases}$$

By part (a) we see that $C(\beta)/\beta$ is decreasing on $[0, \beta_0]$ and continues to decrease for $\beta > \beta_0$. Consequently its supremum is achieved as β approches 0 from above,

$$\sup_{\beta>0} \frac{C(\beta)}{\beta} = \lim_{\beta \searrow 0} \frac{I(\beta)}{\beta}.$$

(d) Note that

$$I(\beta) = \beta \sum_y p(y|1) \log \frac{p(y|1)}{\beta p(y|1) + (1-\beta)p(y|0)} \\ + (1-\beta) \sum_y p(y|0) \log \frac{p(y|0)}{\beta p(y|1) + (1-\beta)p(y|0)}$$

and thus

$$I(\beta)/\beta = \sum_y p(y|1) \log \frac{p(y|1)}{\beta p(y|1) + (1-\beta)p(y|0)} \\ + (1-\beta)\frac{1}{\beta} \sum_y p(y|0) \log \frac{p(y|0)}{\beta p(y|1) + (1-\beta)p(y|0)}$$

As $\beta \to 0$, the first term approaches $\sum_y p(y|1) \log \frac{p(y|1)}{p(y|0)}$. So we only need to show that the second term approaches zero to reach the desired conclusion. To that end, observe that the second term, being in the form of a divergence, is non-negative. Furthermore, using $\ln z \le z - 1$,

$$\begin{aligned} 0 \le \frac{1-\beta}{\beta} & \sum_y p(y|0) \ln \frac{p(y|0)}{\beta p(y|1) + (1-\beta)p(y|0)} \\ & \le \frac{1-\beta}{\beta} \sum_y p(y|0) \Big[\frac{p(y|0)}{\beta p(y|1) + (1-\beta)p(y|0)} - 1 \Big] \\ & = (1-\beta) \sum_y p(y|0) \frac{p(y|0) - p(y|1)}{\beta p(y|1) + (1-\beta)p(y|0)} \\ & \to \sum_y [p(y|0) - p(y|1)] = 0 \quad \text{as } \beta \to 0. \end{aligned}$$

(e) Using $[p(0|0), p(1|0)] = [1-p, p]$ and $[p(0|1), p(1|1)] = [p, 1-p]$ we see that the formula of (d) gives the capacity per cost as $(1-2p) \log[(1-p)/p]$.

Problem 3.

(a) The blocklength is $n = 4$. The number of codewords is 9: we can choose any $x_3 \in \mathbb{F}_3$ and $x_4 \in \mathbb{F}_3$ and the parity check equations determine (x_1, x_2). The rate is $R = (1/4) \log 9 = (1/2) \log 3$.

(b) The received word $\mathbf{y}$ is of the form $\mathbf{x}+\mathbf{z}$ where $\mathbf{z}$ is either the zero vector or it contains an single non-zero entry. Thus $H\mathbf{y}$ is either zero or is a multiple of a column of H. Since all these cases are distinct the decoder can discover if $\mathbf{x}$ was changed during transmission and how.

(c) Augmenting the matrix will increse the blocklength to $n = 5$ and increase the number of codewords to 27: now (x_3, x_4, x_5) can be freely chosen. The rate thus becomes $(3/5)\log 3$ and is increased from $(1/2)\log 3$.

(d) None of them, since each is a multiple of an existing column on H.

(e) In this case the received word is of the form $\mathbf{y} = \mathbf{x} + \mathbf{z}$ where $\mathbf{z}$ is either 0 or contains a single 1. To ensure $H\mathbf{y}$ is distinct in all the cases all we need to ensure is that the columns of H are non-zero and distinct (but not necessarily their multiples). Thus, $\left[\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ are valid augmentations.