ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 14 Information Theory and Coding
Solutions to Homework 6 Oct. 29, 2018

PROBLEM 1. Since L is linear, we know that

$$L(\lambda x) = \lambda L(x)$$

for any $\lambda \in \mathbb{R}$. Similarly, g is concave so it must satisfy the following by definition.

$$g(\lambda x_1 + (1-\lambda)x_2) \geq \lambda g(x_1) + (1-\lambda)g(x_2)$$

for any $\lambda \in [0,1]$. Combining these two statements, the following steps show that f is concave.

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &= g(L(\lambda x_1 + (1-\lambda)x_2)) \\ &= g(\lambda L(x_1) + (1-\lambda)L(x_2)) \qquad (1) \\ &\geq \lambda g(L(x_1)) + (1-\lambda)g(L(x_2)) \qquad (2) \\ &= \lambda f(x_1) + (1-\lambda)f(x_2) \end{aligned}$$

where (1) uses the linearity property of L and (2) uses the concavity property of g.

PROBLEM 2.

(a) Let $s(m) = 0 + 1 + \cdots + (m-1) = m(m-1)/2$. Suppose we have a string of length $n = s(m)$. Then, we can certainly parse it into m words of lengths 0, 1, ..., $(m-1)$, and since these words have different lengths, we are guaranteed to have a distinct parsing. Since a parsing with the maximal number of distinct words will have at least as many words as this particular parsing, we conclude that whenever $n = m(m-1)/2$, $c \geq m$ (and for $n > m(m-1)/2$ we can parse the first $m(m-1)/2$ letters to m, as we just described, and append the remaining letters to the last word to have a parsing into m distinct words).

(b) An all zero string of length $s(m)$ can be parsed into at most m words: in this case distinct words must have distinct lengths and the bound is met with equality.

(c) Now, given n, we can find m such that $s(m-1) \leq n < s(m)$. A string with n letters can be parsed into $m-1$ distinct words by parsing its initial segment of $s(m-1)$ letters with the above procedure, and concatenating the leftover letters to the last word. Thus, if a string can be parsed into $m-1$ distinct words, then $n < s(m)$, and in particular, $n < s(c+1) = c(c+1)/2$. From above, it is clear that no sequence will meet the bound with equality.

PROBLEM 3. Observe that $H(Y) - H(Y|X) = I(X;Y) = I(X;Z) = H(Z) - H(Z|X)$.

(a) Consider a channel with binary input alphabet $\mathcal{X} = \{0,1\}$ with X uniformly distributed over $\mathcal{X}$, output alphabet $\mathcal{Y} = \{0,1,2,3\}$, and probability law

$$P_{Y|X}(y|x) = \begin{cases} \frac{1}{2}, & \text{if } x = 0 \text{ and } y = 0 \\ \frac{1}{2}, & \text{if } x = 0 \text{ and } y = 1 \\ \frac{1}{2}, & \text{if } x = 1 \text{ and } y = 2 \\ \frac{1}{2}, & \text{if } x = 1 \text{ and } y = 3 \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify $H(Y|X) = 1$. Since Y takes any value in $\mathcal{Y}$ with equal probability, its entropy is $H(Y) = 2$. Therefore $I(X;Y) = 1$. Define the processor output to be in alphabet $\mathcal{Z}$ and construct a deterministic processor $g : y \mapsto z = g(y)$ such that,

$$\begin{aligned} g: \quad & \mathcal{Y} \to \mathcal{Z} = \{0, 1\} \\ & 0 \mapsto 0 \\ & 1 \mapsto 0 \\ & 2 \mapsto 1 \\ & 3 \mapsto 1. \end{aligned}$$

Clearly, $H(Z|X) = 0$ and $H(Z) = 1$. Therefore $I(X;Z) = 1$. We conclude that $I(X;Z) = I(X;Y)$ and $H(Z) < H(Y)$.

(b) Consider an error-free channel with binary input alphabet $\mathcal{X} = \{0, 1\}$ with X uniformly distributed over $\mathcal{X}$, binary output alphabet $\mathcal{Y} = \{0, 1\}$, and probability law

$$P_{Y|X}(y|x) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise.} \end{cases}$$

Choose now $\mathcal{Z} = \{0, 1, 2, 3\}$ an construct a probabilistic processor G such that

$$\begin{aligned} G: \quad & \mathcal{Y} \to \mathcal{Z} \\ & 0 \mapsto 0 \ \text{ with probability } \frac{1}{2} \ \text{ or } \ 1 \ \text{ with probability } \frac{1}{2} \\ & 1 \mapsto 2 \ \text{ with probability } \frac{1}{2} \ \text{ or } \ 3 \ \text{ with probability } \frac{1}{2}. \end{aligned}$$

Clearly, $I(X;Y) = 1 = I(X;Z)$ and $H(Y) = 1 < 2 = H(Z)$.

Problem 4.

(a)

$$\Pr(U = u|V = ?) = \frac{\Pr(V = ?|U = u)\, p_U(u)}{\Pr(V = ?)} = \frac{p_U(u)p}{p} = p_U(u)$$

(b)

$$\begin{aligned} I(U;V) &= H(U) - H(U|V) \\ &= H(U) - \Pr(V = ?)H(U|V = ?) - \Pr(V \neq ?)H(U|V \neq ?) \\ &\stackrel{(a)}{=} H(U) - p \sum_{u=1}^{K} \Pr(U = u|V = ?) \log \frac{1}{\Pr(U = u|V = ?)} \\ &\stackrel{(b)}{=} H(U) - p \sum_{u=1}^{K} p_U(u) \log \frac{1}{p_U(u)} = H(U) - pH(U) = (1 - p)H(U), \end{aligned}$$

where (a) is obtained by noticing that if $V \neq ?$ then $V = U$ and $H(U|V \neq ?) = 0$ and (b) is obtained since $\Pr(U = u|V = ?) = p_U(u)$.

(c) Let C_p be the capacity of this channel. Then,

$$C_p = \max_{p_U} I(U,V) = \max_{p_U}(1-p)H(U) = (1-p)\max_{p_U} H(U) = (1-p)\log K,$$

with the maximum achieved when U is uniformly distributed over $\{1, \cdots, K\}$.

Problem 5.

(a) Since the channel is symmetric, the input distribution that maximizes the mutual information is the uniform one. Therefore, $C = 1 + \epsilon \log_2(\epsilon) + (1-\epsilon)\log_2(\epsilon)$ which is equal to 0 when $\epsilon = \frac{1}{2}$.

(b) We have

- $I(X^n; Y^n) = I(X_2^n; Y^{n-1}) + I(X_2^n; Y_n | Y^{n-1}) + I(X_1; Y^n | X_2^n)$.
- $X_2^n = Y^{n-1}$ and $Y_1, \ldots, Y_n$ are i.i.d. and uniform in $\{0,1\}$, so $I(X_2^n; Y^{n-1}) = H(Y^{n-1}) = n-1$.
- Y_n is independent of (X_2^n, Y^{n-1}), so $I(X_2^n; Y_n | Y^{n-1}) = 0$.
- X_1 is independent of (Y^n, X_2^n), so $I(X_1; Y^n | X_2^n) = 0$.

Therefore, $I(X^n; Y^n) = n-1$.

(c) W is independent of Y^n, so $I(W; Y^n) = 0 = nC$.