

ECE1502S — Information Theory Midterm Test Solution

1. (*Keeping Both Eyes Open*) We know that X and Y are independent and X and Z are independent. Does this mean that X is independent of the pair Y, Z ? The answer is no! For example, let Y and Z be chosen independently and uniformly at random from $\{0, 1\}$, and let $X = Y \oplus Z$ where \oplus denotes modulo-two addition. Then X is independent of Y and X is independent of Z , but $I(X; Y, Z) = 1$ bit.
2. (*Entropy Inequalities*) In general, we have $H(X_2|X_1) \leq \frac{1}{2}H(X_1, X_2) \leq H(X_1)$. To see that $2H(X_2|X_1) \leq H(X_1, X_2)$ observe that

$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2|X_1) \text{ (chain rule)} \\ &= H(X_2) + H(X_2|X_1) \text{ (by stationarity)} \\ &\geq H(X_2|X_1) + H(X_2|X_1) \text{ (conditioning does not increase entropy)} \\ &= 2H(X_2|X_1) \end{aligned}$$

Similarly,

$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2|X_1) \text{ (chain rule)} \\ &\leq H(X_1) + H(X_2) \text{ (conditioning does not increase entropy)} \\ &= H(X_1) + H(X_1) \text{ (by stationarity)} \\ &= 2H(X_1). \end{aligned}$$

(More cleverly, we may observe that $H(X_1, X_2) = H(X_1) + H(X_2|X_1)$, where clearly $H(X_2|X_1) \leq H(X_2) = H(X_1)$. Thus the numbers $H(X_2|X_1)$, $\frac{1}{2}H(X_1, X_2)$ and $H(X_1)$ form a non-decreasing arithmetic sequence, and the result follows.)

3. (*Lempel-Ziv Parsing*)

(a) The parsing continues as indicated by commas:

0, 00, 001, 01, 1, 11, 0010, 10, 101.

(b) Any sequence consisting of all sequences of length one, followed by all sequences of length two, followed by all sequences of length three, etc., will do; e.g.,

0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, 0001, ...

The first two phrases have length one, the next four have length two, the next eight have length three, etc. Stated another way, the first 2 phrases have length ≤ 1 , the first 6 phrases have length ≤ 2 , the first 14 phrases have length ≤ 3 , etc. In general, the first $2(2^m - 1) = 2^{m+1} - 2$ phrases have length $\leq m$. In particular, the first 62 phrases have length ≤ 5 and the first 126 phrases phrases have length ≤ 6 . Thus the 100th phrase has length 6.

(c) A constant sequence will do, e.g.,

$$0, 00, 000, 0000, 00000, 000000, 0000000, \dots$$

The i th phrase has length i ; in particular, the 100th phrase has length 100.

4. (*Entropy Bounds*) Let $\mathcal{S}_X = \{x_1, \dots, x_m\}$ with $P[x_1] = p_1$. Define a new variable Z that takes value 1 if $X = x_1$ and takes value 0 otherwise. Then

$$H(X, Z) = H(X) + H(Z|X) = H(X),$$

which follows from the fact that Z is a deterministic function of X and so $H(Z|X) = 0$. On the other hand

$$\begin{aligned} H(X) = H(X, Z) &= H(Z) + H(X|Z) \\ &= H(Z) + p_1 H(X|Z=1) + (1-p_1) H(X|Z=0) \\ &= H(Z) + (1-p_1) H(X|Z=0), \end{aligned}$$

where the latter equality follows from the fact that $H(X|Z=1) = 0$ (if $Z = 1$, there is no uncertainty in the value of X). Note that if $Z = 0$, then $X \neq x_1$, so X is effectively distributed over $\{x_2, \dots, x_m\}$, an alphabet of $m-1$ symbols. It follows that $0 \leq H(X|Z=0) \leq \log_2(m-1)$, where the lower bound is achieved by any distribution having all of its mass $(1-p_1)$ on a single element of $\{x_2, \dots, x_m\}$ and the upper bound is achieved by any distribution having its mass uniformly distributed over $\{x_2, \dots, x_m\}$, i.e., assigning $(1-p_1)/(m-1)$ mass to each. Since $H(Z) = \mathcal{H}(p_1)$ we obtain the bounds

$$\mathcal{H}(p_1) \leq H(X) \leq \mathcal{H}(p_1) + (1-p_1) \log_2(m-1)$$

In particular, when $p_1 = 1/2$, we get

$$1 \leq H(X) \leq 1 + \frac{1}{2} \log_2(m-1).$$

5. (*Perfect Security*) We are told that (a) $I(X;K) = 0$, (b) $I(X;Y) = 0$ and (c) $H(X|Y, K) = 0$.

From the chain rule we have

$$\begin{aligned} H(X, K, Y) &= H(X) + \underbrace{H(K|X)}_{H(K)} + \underbrace{H(Y|X, K)}_0 \\ &= H(X) + H(K) \end{aligned} \tag{1}$$

We also have

$$\begin{aligned} H(X, K, Y) &= H(K) + H(Y|K) + \underbrace{H(X|Y, K)}_0 \\ &= H(K) + H(Y|K) \end{aligned} \tag{2}$$

From (1) and (2) it follows that $H(Y|K) = H(X)$. Since $H(Y) \geq H(Y|K)$ (conditioning does not increase entropy) we obtain the result that $H(Y) \geq H(X)$, i.e., Y should have at least as much entropy as X , and so $m \geq n$.

We also have

$$\begin{aligned} H(X, K, Y) &= H(X) + \underbrace{H(Y|X)}_{H(Y)} + H(K|Y, X) \\ &= H(X) + H(Y) + H(K|Y, X). \end{aligned} \tag{3}$$

Combining (1) and (3) we find that $H(K) = H(Y) + H(K|Y, X)$ and so certainly $H(K) \geq H(Y)$. But since $H(Y) \geq H(X)$, we have $H(K) \geq H(X)$, and so $k \geq n$.

We can achieve equality ($k = m = n$) via a one-time pad: choose K uniformly in $\{0, 1\}^n$ and let $Y = X \oplus K$, where \oplus denotes a modulo-two addition (XOR). Then Y is independent of X yet X can be recovered from Y and K as $X = Y \oplus K$.

6. (*Matching with Dyadic Distributions*) Let $q_i = 2^{-l_i}$ where l_i is an integer. We have

$$\begin{aligned} D(p||q) &= \sum_{i=1}^m p_i \log_2(p_i/q_i) \\ &= \sum_{i=1}^m p_i \log_2(2^{l_i} p_i) \\ &= \sum_{i=1}^m p_i l_i + \sum_{i=1}^m p_i \log_2 p_i \\ &= \sum_{i=1}^m p_i l_i - H(X) \end{aligned}$$

To minimize $D(p||q)$ we must minimize $\sum_i p_i l_i$ subject to $\sum_i 2^{-l_i} = 1$ and $l_i \in \{1, 2, 3, \dots\}$. This is the same problem as finding an optimal prefix code that achieves the Kraft inequality with equality.

To solve this problem, we can use the Huffman procedure to obtain an optimal code C . If the codeword assigned to outcome i has length l_i , then $q(i) = 2^{-l_i}$. Since dummy symbols are never required in the binary case, the Huffman code will satisfy the Kraft inequality with equality, i.e., $\sum_i 2^{-l_i} = 1$. Since the Huffman code is not unique, neither is the best dyadic approximation to a given distribution. For example if $m = 4$ and $(p_1, \dots, p_4) = (1/3, 1/3, 1/6, 1/6)$, then the dyadic distributions $(1/4, 1/4, 1/4, 1/4)$ and $(1/2, 1/4, 1/8, 1/8)$ both minimize $D(p||q)$.