ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 7 Information Theory and Coding
Solutions to Homework 3 Oct. 6, 2015

Problem 1.

(a) Since the lengths of the codewords satisfy the Kraft inequality, an instantaneous code can be used for the final stage of encoding the intermediate digits to binary codewords. In this case, each stage of the encoding is uniquely decodable, and thus the overall code is uniquely decodable.

(b) The indicated source sequences have probabilities $0.1, (0.9)(0.1), (0.9)^2(0.1), (0.9)^3(0.1), \ldots, (0.9)^7(0.1), (0.9)^8$. Thus,

$$\bar{N} = \sum_{i=1}^{8} i(0.1)(0.9)^{i-1} + 8(0.9)^8 = 5.6953.$$

(c)

$$\bar{M} = 1(0.9)^8 + 4[1-(0.9)^8] = 2.7086.$$

(d) Let $N(i)$ be the number of source digits giving rise to the first i intermediate digits. For any $\epsilon > 0$

$$\lim_{i\to\infty} \Pr\Big[\Big|\frac{N(i)}{i} - \bar{N}\Big| > \epsilon\Big] = 0.$$

Similarly, let $M(i)$ be the number of encoded bits corresponding the the first i intermediate digits. Then

$$\lim_{i\to\infty} \Pr\Big[\Big|\frac{M(i)}{i} - \bar{M}\Big| > \epsilon\Big] = 0.$$

From this, we see that for any $\epsilon > 0$,

$$\lim_{i\to\infty} \Pr\Big[\Big|\frac{M(i)}{N(i)} - \frac{\bar{M}}{\bar{N}}\Big| > \epsilon\Big] = 0,$$

and that for a long source sequence the number of encoded bits per source digit will be $\bar{M}/\bar{N} = 0.4756$.

The average length of the Huffman code encoding 4 source digits at a time is 1.9702, yielding $1.9702/4 = 0.49255$ encoded bits per source digit.

For those of you puzzled by the fact that the 'optimum' Huffman code gives a worse result for this source than the run-length coding technique, observe that the Huffman code is the optimal solution to a mathematical problem with a given message set, but the choice of a message set can be more important than the choice of codewords for a given message set.

Problem 2.

(a) $H(X) = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3 = 0.918 \text{ bits} = H(Y)$.

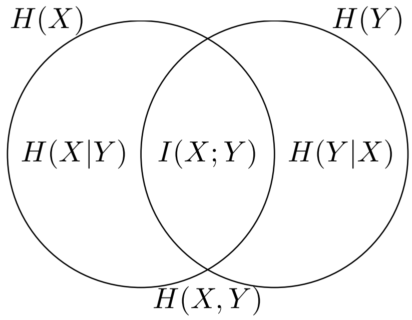
(b) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = 0.667 \text{ bits} = H(Y|X)$.

(c) $H(X,Y) = 3 \times \frac{1}{3}\log 3 = 1.585$ bits.

(d) $H(Y) - H(Y|X) = 0.251$ bits.

(d) $I(X;Y) = H(Y) - H(Y|X) = 0.251$ bits.

(f)



Problem 3.

$$\begin{aligned} H(X) &= -\sum_{k=1}^{M} P_X(a_k)\log P_X(a_k) \\ &= -\sum_{k=1}^{M-1}(1-\alpha)P_Y(a_k)\log[(1-\alpha)P_Y(a_k)] - \alpha\log\alpha \\ &= (1-\alpha)H(Y) - (1-\alpha)\log(1-\alpha) - \alpha\log\alpha \end{aligned}$$

Since Y is a random variable that takes $M-1$ values $H(Y) \le \log(M-1)$ with equality if and only if Y takes each of its possible values with equal probability.

Problem 4.

(a) Using the chain rule for mutual information,

$$I(X,Y;Z) = I(X;Z) + I(Y;Z \mid X) \ge I(X;Z),$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

(b) Using the chain rule for conditional entropy,

$$H(X,Y \mid Z) = H(X \mid Z) + H(Y \mid X,Z) \ge H(X \mid Z),$$

with equality iff $H(Y \mid X,Z) = 0$, that is, when Y is a function of X and/or Z.

(c) Using first the chain rule for entropy and then the definition of conditional mutual information,

$$\begin{aligned} H(X,Y,Z) - H(X,Y) = H(Z \mid X,Y) &= H(Z \mid X) - I(Y;Z \mid X) \\ &\le H(Z \mid X) = H(X,Z) - H(X), \end{aligned}$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

(d) Using the chain rule for mutual information,

$$I(X;Z \mid Y) + I(Z;Y) = I(X,Y;Z) = I(Z;Y \mid X) + I(X;Z)\,,$$

and therefore

$$I(X;Z \mid Y) = I(Z;Y \mid X) - I(Z;Y) + I(X;Z)\,.$$

We see that this inequality is actually an equality in all cases.

Problem 5. Let X^i denote $X_1, \ldots, X_i$.

(a) By stationarity we have for all $1 \leq i \leq n$,

$$H(X_n|X^{n-1}) \leq H(X_n|X_{n-i+1}, X_{n-i+2}, \ldots, X_{n-1}) = H(X_i|X^{i-1}),$$

which implies that,

$$\begin{aligned} H(X_n|X^{n-1}) &= \frac{\sum_{i=1}^n H(X_n|X^{n-1})}{n} && (1)\\ &\leq \frac{\sum_{i=1}^n H(X_i|X^{i-1})}{n} && (2)\\ &= \frac{H(X_1, X_2, \ldots, X_n)}{n}. && (3) \end{aligned}$$

(b) By the chain rule for entropy,

$$\begin{aligned} \frac{H(X_1, X_2, \ldots, X_n)}{n} &= \frac{\sum_{i=1}^n H(X_i|X^{i-1})}{n} && (4)\\ &= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n} && (5)\\ &= \frac{H(X_n|X^{n-1}) + H(X_1, X_2, \ldots, X_{n-1})}{n}. && (6) \end{aligned}$$

From stationarity it follows that for all $1 \leq i \leq n$,

$$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

which further implies, by summing both sides over $i = 1, \ldots, n-1$ and dividing by $n-1$, that,

$$\begin{aligned} H(X_n|X^{n-1}) &\leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} && (7)\\ &= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. && (8) \end{aligned}$$

Combining (6) and (8) yields,

$$\begin{aligned} \frac{H(X_1, X_2, \ldots, X_n)}{n} &\leq \frac{1}{n}\left[\frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1} + H(X_1, X_2, \ldots, X_{n-1})\right] && (9)\\ &= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. && (10) \end{aligned}$$

PROBLEM 6. By the chain rule for entropy,

$$
\begin{aligned}
H(X_0|X_{-1},\ldots,X_{-n}) &= H(X_0,X_{-1},\ldots,X_{-n}) - H(X_{-1},\ldots,X_{-n}) &(11)\\
&= H(X_0,X_1,\ldots,X_n) - H(X_1,\ldots,X_n) &(12)\\
&= H(X_0|X_1,\ldots,X_n), &(13)
\end{aligned}
$$

where (12) follows from stationarity.

PROBLEM 7. For a Markov chain, X_0 and X_n are independent given X_{n-1}. Thus

$$H(X_0|X_nX_{n-1}) = H(X_0|X_{n-1})$$

But, since conditioning reduces entropy,

$$H(X_0|X_nX_{n-1}) \leq H(X_0|X_n).$$

Putting the above together we see that $H(X_0|X_{n-1}) \leq H(X_0|X_n)$.