# 3F7 Information Theory and Coding
# Engineering Tripos 2017 – Solutions

**Question** 1

(a) i) $P(X=1)=5/16,\ P(X=2)=3/8,\ P(X=2)=5/16$. This can be used to compute:

$$H(X) = \frac{2\ .\ 5}{16}\log_2\frac{16}{5} + \frac{3}{8}\log_2\frac{8}{3} = 1.5794 \text{ bits}.$$

The joint entropy is

$$H(X,Y) = \sum_{x,y} P_{XY}(x,y)\log_2\frac{1}{P_{XY}(x,y)} = \frac{4}{16}\log_2(16) + \frac{3}{4}\log_2(4) = 1 + 1.5 = 2.5 \text{ bits}.$$

The conditional entropy is $\quad H(Y|X) = H(X,Y) - H(X) = 0.9206$ bit. [20%]

ii) The pmf of $Z$ is

| $z$ | 1 | 2 | 3 | 4 | 6 | 9 |
|---|---|---|---|---|---|---|
| $P(z)$ | 1/4 | 1/8 | 0 | 1/4 | 1/8 | 1/4, |

From this we can compute: [15%]

$$H(Z) = \sum_z P(z)\log_2 1/P(z) = \frac{2}{8}\log_2(8) + \frac{3}{4}\log_2(4) = 0.75 + 1.5 = 2.25 \text{ bits}.$$

iii) We have

$$\begin{aligned} H(Z|X) &= P(X=1)H(Z|X=1) + P(X=2)H(Z|X=2) + P(X=3)H(Z|X=3) \\ &= P(X=1)H(Y|X=1) + P(X=2)H(Y|X=2) + P(X=3)H(Y|X=3) \\ &= H(Y|X) = 0.9206\text{bit, from part a.(i)} \end{aligned}$$

Next,

$$\begin{aligned} & H(X,Z) = H(X) + H(Z|X) = H(Z) + H(X|Z) \\ \Rightarrow\ & H(X|Z) = H(X) + H(Z|X) - H(Z) = 1.5794 + 0.9206 - 2.25 = 0.25 \text{ bit} \end{aligned}$$

[20%]

(b) i) The exponential density, $f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}, \quad x \geq 0$, has unit integral and mean $\mu$:

$$\begin{aligned} \text{therefore}\quad h(f) &= -\int_0^\infty f(x)\log_2 f(x)dx = \int_0^\infty (-\log_2\frac{1}{\mu} + \frac{x}{\mu}\log_2 e)\frac{1}{\mu}e^{-\frac{x}{\mu}}\,dx \\ &= -\log_2\frac{1}{\mu} + \log_2 e\ .\ \frac{1}{\mu}\int_0^\infty \frac{x}{\mu}e^{-\frac{x}{\mu}}\,dx = -\log_2\frac{1}{\mu} + (\log_2 e)\ .\ \frac{\mu}{\mu} \\ &= \log_2\mu + \log_2 e = \log_2(\mu e) \text{ bits}, \\ \text{or}\quad &= \frac{1}{\ln 2}(\ln\mu + 1) \text{ bits}. \end{aligned}$$

[20%]

ii) Let the density of $Y$ be $g$, and let $f$ be the exponential density given in part (b).(i). Consider the relative entropy between $g$ and $f$. We have

$$\begin{aligned} D(g||f) &= \int_{-\infty}^{\infty} g(x) \log_2 \frac{g(x)}{f(x)}\, dx \\ &= \int_0^{\infty} g(x) \log_2 g(x)\, dx + \int_0^{\infty} g(x) \left[\log_2 \frac{1}{\frac{1}{\mu} e^{-\frac{x}{\mu}}}\right] dx \\ &\overset{(a)}{=} -h(Y) \; + \int_0^{\infty} g(x) \left[\frac{x}{\mu} \log_2 e + \log_2 \mu\right] dx \\ &\overset{(b)}{=} -h(Y) \; + \frac{\mu}{\mu} . \log_2 e + 1 . \log_2 \mu \\ &= \log_2(\mu e) - h(Y). \end{aligned} \tag{1}$$

Step $(a)$ is obtained from the definition of differential entropy $h(Y)$; step $(b)$ holds because the mean of $Y$ (with density $g$) is $\mu$, and the fact that $g$ being a density integrates to 1.

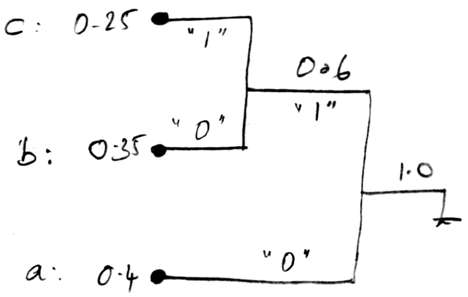
As the relative entropy is non-negative, we have from (1)

$$D(g||f) = \log_2(\mu e) - h(Y) \geq 0 \quad \Rightarrow \quad h(Y) \leq \log_2(\mu e) = h(X),$$

where the last equality was shown in part (b)(i) above. [25%]

## Question 2

(a) i) An optimal prefix-free code can be found using the Huffman procedure as shown below.




| Symbol | Codeword |
|---|---|
| a | 0 |
| b | 10 |
| c | 11 |

(Other codewords are also possible, by interchanging 0 and 1.) The expected code length is [15%]

$$L = (0.4)\ .\ 1\ +\ (0.35)\ .\ 2\ +\ (0.25)\ .\ 2 = 1.6 \text{ bits}$$

ii) The minimum expected length in bits/symbol of any uniquely decodable code is the entropy $H(X)$, which equals [5%]

$$-0.4 \log_2(0.4) - 0.35 \log_2(0.35) - 0.25 \log_2(0.25) = 1.5589 \text{ bits/symbol}$$

iii) A practical technique for prefix-free coding of long sequences is arithmetic coding, which proceeds as follows. Start by dividing the interval $[0, 1)$ in the proportion of the symbol probabilities, e.g, $[0, .4)$, $[0.4, .75)$, $[.75, 1)$. Choose the sub-interval corresponding to the first symbol $X_1$, and divide the chosen subinterval into three further sub-intervals in the proportion of the symbol probabilities. Then choose the second-level subinterval corresponding to $X_2$ and so on. Finally, after finding the subinterval corresponding to $X_1, \ldots, X_n$, find a dyadic interval of the form $\left[\frac{j}{2^\ell}, \frac{j+1}{2^\ell}\right)$ that lies inside this sub-interval. The binary representation of $j$ is the binary codeword.

The length of the binary codeword for $X_1, \ldots, X_n$ is at most $\lceil \log_2 \frac{1}{P(X_1, \ldots, X_n)} \rceil + 1$ which gives an expected code length [20%]

$$L = \frac{1}{n} \sum_{x^n} P(x^n) \left( \lceil \log_2 \frac{1}{P(X_1, \ldots, X_n)} \rceil + 1 \right) < H(X) + \frac{2}{n} \text{ bits/symbol.}$$

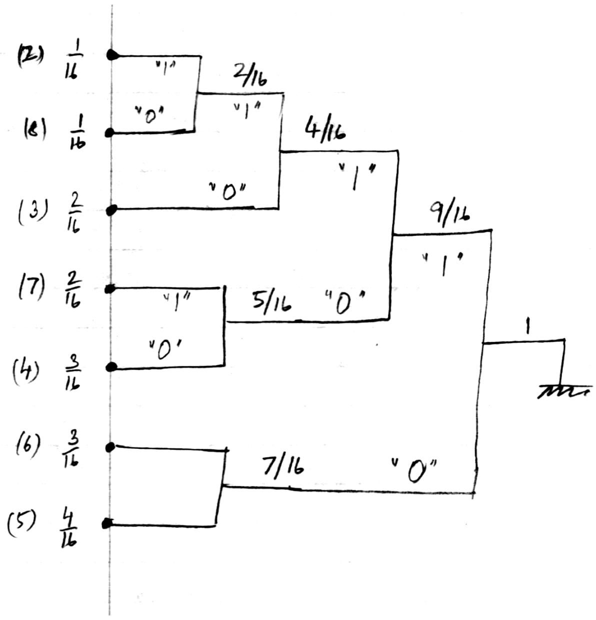
(b) i) Note that $Z$ can take values in $\{2, \ldots, 8\}$ with

$$P_Z(2) = \frac{1}{16},\ P_Z(3) = \frac{2}{16},\ P_Z(4) = \frac{3}{16},\ P_Z(5) = \frac{4}{16},\ P_Z(6) = \frac{3}{16},\ P_Z(7) = \frac{2}{16},\ P_Z(8) = \frac{1}{16}.$$

The expected number of questions with the given strategy is [25%]

$$\begin{aligned} L_1 &= 1P_Z(2) + 2P_Z(3) + 3P_Z(4) + 4P_Z(5) + 5P_Z(6) + 6P_Z(7) + 6P_Z(8) \\ &= \frac{1}{16}[1 + 4 + 9 + 16 + 15 + 12 + 6] = \frac{63}{16} = 3.9375 \end{aligned}$$

ii) An optimal strategy can be found using the Huffman procedure, as shown below.




| Symbol | Codeword |
|---|---|
| 5 | 0 0 |
| 6 | 0 1 |
| 4 | 1 0 0 |
| 7 | 1 0 1 |
| 3 | 1 1 0 |
| 8 | 1 1 1 0 |
| 2 | 1 1 1 1 |

(Note that there are many valid answers here, but all correct answers will give the same expected number of questions in (b)(iii) below.)

From the root of the tree, the questions will be of the form:
– Is the value either 5 or 6: If 'yes', follow the lower '0' branch in the tree; if 'no', follow the upper '1' branch.
– Repeat the above procedure with an appropriate new question until a leaf node is reached. [20%]

iii) The expected code length (and hence number of questions) for the optimal strategy derived above is

$$\begin{aligned} L_{opt} &= 2[P_Z(5) + P_Z(6)] + 3[P_Z(4) + P_Z(7) + P_Z(3)] + 4[P_Z(8) + P_Z(2)] \\ &= \frac{1}{16}[2\,.\,7 + 3\,.\,7 + 4\,.\,2] = \frac{43}{16} = 2.6875 \end{aligned}$$

With the above strategy, the sequence of questions asked if the value is 3 is:

– Is the value either 5 or 6 (Ans: No)

– Is the value either 4 or 7 (Ans: No)

– Is the value 3 (Ans: Yes) [15%]

## Question 3

(a) i) $P(Y=0) = \frac{3}{4}p + \frac{1}{2}(1-p) = \frac{1}{2} + \frac{p}{4}$;
$P(Y=1) = \frac{1}{2} - \frac{p}{4}$. [10%]

ii) Let the input distribution be $P(X=0) = p$, $P(X=1) = (1-p)$, as above. Then with $P(Y)$ as computed above, have

$$\begin{aligned} I(X;Y) = H(Y) - H(Y|X) &= H_2\left(\frac{1}{2} + \frac{p}{4}\right) - p\ H(Y|X=0) - (1-p)\ H(Y|X=1) \\ &= H_2\left(\frac{1}{2} + \frac{p}{4}\right) - p \cdot H_2(0.75) - (1-p) \cdot H_2(0.5) \qquad (2) \\ &= H_2\left(\frac{1}{2} + \frac{p}{4}\right) - 1 + p \cdot (1 - H_2(0.75)) \end{aligned}$$

To find the capacity, we need to maximize $I(X;Y)$ over $p$. Using the hint to differentiate we find

$$\frac{d}{dp} I(X;Y) = \frac{1}{4} \log_2 \left( \frac{\frac{1}{2} - \frac{p}{4}}{\frac{1}{2} + \frac{p}{4}} \right) + 1 - H_2(0.75) \qquad (3)$$

Setting (3) to 0 and solving for $p$ yields the optimum value $p^*$, where:

$$\frac{2-p^*}{2+p^*} = 2^{4(H_2(0.75)-1)} = 2^{3\log_2(4/3) + \log_2(4) - 4} = (4/3)^3 \cdot 4/16 = \frac{16}{27}$$

Hence $p^* = \dfrac{2(1-16/27)}{1+16/27} = \dfrac{2(27-16)}{(27+16)} = \dfrac{22}{43} = 0.5116$

Substituting this value into (2), we get the capacity:

$$\mathcal{C} = I(X;Y)_{\max} = 0.9523 - 1 + 0.5116(1 - 0.8113) = 0.0488 \text{ bits}$$

and the maximising input distribution is $P(X) = \{p^*, 1-p^*\} = \{0.5116, 0.4884\}$. [40%]

iii) Fix a rate $R$ just less than $\mathcal{C}$ and a sufficiently large block length $n$. Construct a codebook of $2^{nR}$ length-$n$ codewords with each symbol of each codeword generated randomly and i.i.d according to $P_X(0) = 0.5116, P_X(1) = 0.4884$. A joint typicality decoder can be used to recover the transmitted codeword with arbitrarily low probability of error by taking sufficiently large $n$. [20%]

(b) Step $(a)$ is obtained by using the chain rule for conditional entropy:

$$I(X^n;Y^n) \overset{(a)}{=} H(Y^n) - H(Y^n|X^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_{i-1}, \ldots, Y_1, X^n)$$

Step $(b)$ holds because the channel is memoryless, so given all the inputs $(X_1, \ldots, X_n)$ and the past outputs $(Y_1, \ldots, Y_{i-1})$, the current output $Y_i$ depends *only* on the current input $X_i$, so that for each $i$:

$$H(Y_i \mid Y_{i-1}, \ldots, Y_1, X^n) = H(Y_i \mid X_i).$$

Step $(c)$ is obtained first by using the chain rule for entropy, and then the fact that removing the conditioning can only increase the entropy (or equivalently, adding conditioning can only decrease the entropy):

$$\begin{aligned} H(Y^n) &= H(Y_1) + H(Y_2|Y_1) + \ldots + H(Y_n|Y_{n-1}, \ldots, Y_1) \\ &\leq H(Y_1) + H(Y_2) + \ldots + H(Y_n) \end{aligned}$$

Step $(d)$ is obtained by observing that, for each $i$, $H(Y_i) - H(Y_i|X_i) = I(X_i;Y_i)$ and $I(X_i;Y_i) \leq \mathcal{C}$, since $\mathcal{C} = \max(I(X;Y))$ over all input distributions. [30%]

## Question 4

(a) i) Dimension $k = 3$, block length $n = 7$, rate $R = 3/7$. [10%]

ii) Using the data sheet to get the parity check matrix for a given generator matrix, we get

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

[10%]

iii) We solve for $\underline{c} = [c_1, c_2, c_3, 1, c_5, 0, 0]$ using $\underline{c}\mathbf{H}^T = \underline{0}$:

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ 1 \\ c_5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This gives the equations: [25%]

$$c_2 + c_3 + 1 = 0 \qquad c_1 + c_3 + c_5 = 0 \qquad c_1 + c_2 = 0 \qquad c_1 + c_2 + c_3 = 0$$

Solving these gives $c_1 = 1,\ c_2 = 1,\ c_3 = 0,\ c_5 = 1$. Hence $\underline{c} = [1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 0]$.

As this is a systematic code, the information bits are $[x_1 = c_1,\ x_2 = c_2,\ x_3 = c_3] = [1,\ 1,\ 0]$.

(One can also solve the problem starting from the equations obtained from $[x_1, x_2, x_3]\mathbf{G} = \underline{c}$, using just the non-erased bits for $\underline{c}$ and the corresponding columns from $\mathbf{G}$.)

(b) i) From the Data Sheet, the design rate of the code is given by:

$$R = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx} = 1 - \frac{\left[x^6/6\right]_0^1}{\left[0.1x^3 + 0.08x^5 + 0.05x^6\right]_0^1} = \frac{19}{69} = 0.2754$$

[10%]

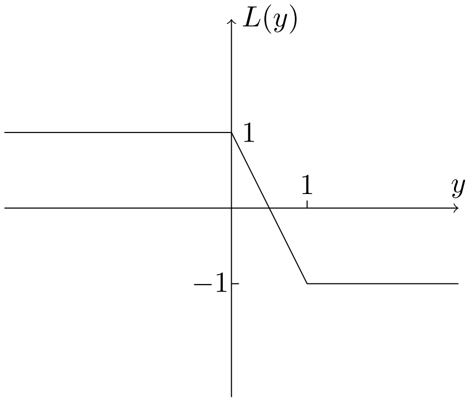
ii) The given $\lambda(x)$ tells us that 30% of edges are connected to degree-3 variable nodes, 40% of edges are connected to degree-5 variable nodes, and 30% of edges are connected to degree-6 variable nodes. [10%]

$\rho(x)$ tells us that all the edges are connected to degree-6 check nodes, i.e. all check nodes have degree 6.

iii) The log-likelihood ratio is [20%]

$$\begin{aligned} L(y) &= \ln\left(\frac{f_{Y|X}(y \mid 0)}{f_{Y|X}(y \mid 1)}\right) \\ &= \ln\left(\frac{\exp(-|y|)}{\exp(-|y-1|)}\right) = |y-1| - |y| \\ &= \begin{cases} -(y-1) - (-y) = 1 & \text{for } y < 0, \\ -(y-1) - y = 1 - 2y & \text{for } 0 \le y < 1, \\ (y-1) - y = -1 & \text{for } y \ge 1, \end{cases} \end{aligned}$$

as plotted below:




iv) From part (iii), the log-likelihood ratio corresponding to the channel output $y_j = 0.1$ is $L(y_j) = 1 - 2(0.1) = 0.8$. According to the belief propagation update equation for the variable node in the Data Sheet, the outgoing message sent by the variable node along the third edge should therefore be [15%]

$$L_{ji} = L(y_j) + \sum_{i' \backslash i} L_{i'j} = 0.8 + (-0.23) + 0.53 = +1.1$$

N G Kingsbury
24 May 2017.