3F7 Information Theory and Coding
Engineering Tripos 2020/21 – Solutions

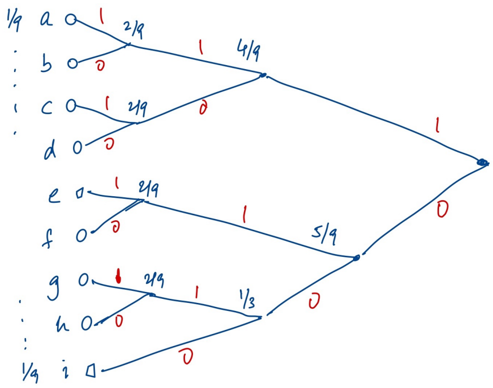
Question 1

(a) Since $X - Y - Z$ form a Markov chain, by the data processing inequality we have [20%]

$$I(X;Z) \leq I(X;Y) = H(Y) - H(Y|X) \leq H(Y).$$

Since Y takes values in a set of cardinality k, we have $H(Y) \leq \log_2 k$. Hence $I(X;Z) \leq \log_2 k$.

(b) i) An optimal binary code can be derived using the Huffman procedure, shown below: [25%]



The codewords are:

$$a \to 111, \quad b \to 110, \quad c \to 101, \quad d \to 100,$$
$$e \to 011, \quad f \to 010, \quad g \to 0011, \quad h \to 0010, \quad i \to 000.$$

The expected codelength is $\frac{7}{9} \cdot 3 + \frac{2}{9} \cdot 4 = 29/9 = 3.22$ bits/symbol.

ii) We have $H(X_k \mid X_k - 1) = \sum_{x \in \{\texttt{a, ..., i}\}} P(X_{k-1} = x) H(X_k \mid X_{k-1} = x)$. To compute this, the probability distributions are $P(X_{k-1} = x) = \frac{1}{9}$, and [25%]

$$P(X_k = x \mid X_{k-1} = x) = 0.99, \quad P(X_k = y \mid X_{k-1} = x) = \frac{0.01}{8} \text{ for } y \neq x, y \in \{\texttt{a, ..., i}\}.$$

Using thus, we compute the conditional entropy:

$$H(X_k \mid X_{k-1}) = \sum_{x \in \{\texttt{a, ..., i}\}} \frac{1}{9} \left[0.99 \log_2 \frac{1}{0.99} + 8 \times \frac{0.01}{8} \log_2 \frac{8}{0.01} \right] = H_2(0.01) + 0.03 = 0.11.$$

iii) Using the chain rule, the joint entropy is

$$\begin{aligned} H(X_1, \ldots, X_N) &= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_2, X_1) + \ldots + H(X_N \mid X_{N-1}, \ldots, X_1) \\ &= H(X_1) + \sum_{k=2}^{N} H(X_k \mid X_{k-1}) \\ &= \log_2 9 + (N-1)(0.11) = (0.11)N + 3.06 \end{aligned}$$

The second equality above holds because $H(X_k \mid X_{k-1}, \ldots X_1) = H(X_k \mid X_{k-1})$, for $k \geq 2$. [20%]

v) Arithmetic coding. Bound on the expected code length for compressing N source symbols: $= H(X_1, \ldots, X_N) + 2 = (0.11)N + 5.06$ bits. [10%]

Assessor's comment: Generally well answered, though many made mistakes in computing the conditional entropy in part (b).(ii).

Question 2

(a) The entropy $H(X) = 0.1614$, and the probability of any sequence x^n is $(0.98)^{n_a}(0.01)^{n_b+n_c}$, where n_a, n_b, n_c are the number of occurrences of a, b, c in the sequence. Therefore

$$\begin{aligned} A_{\epsilon,n} &= \left\{ x^n \, ,: \; 0.16 - \epsilon \, \leq \, \frac{n_a}{n} \log_2 \frac{1}{0.98} + \frac{n_a + n_b}{n} \log_2 \frac{1}{0.01} \, \leq \, 0.16 + \epsilon \right\} \\ &= \left\{ x^n \, ,: \; 0.15 \, \leq \, 0.03 \frac{n_a}{n} + 6.64 \frac{n_a + n_b}{n} \, \leq \, 0.17 \right\} \end{aligned} \tag{1}$$

The length-100 sequences that are in this set are those with $n_a = 98$ and $n_b + n_c = 2$, i.e., there are only three kinds of sequences: 1. 98 a's and 2b's; 2. 98 a's and 2 c's; 3. 98 a's, 1 b and 1 c. [30%]

ii) The total number of such sequences is $\binom{100}{2} 2^2 = 19,800$. Since $\log_2(19800) = 14.27$, the number of bits needed for each codeword is 15. [15%]

(b) Let Q_X be any other pmf over $\{1, \dots, N\}$ with $\sum_{i=1}^{N} i \, Q_X(i) = \alpha$. Since the relative entropy $D(Q_X \| P_X^*)$ is non-negative, we have:

$$\begin{aligned} 0 \leq D(Q_X \| P_X^*) &= \sum_i Q_X(i) \log_2 \frac{Q_X(i)}{2^{-(\lambda_0 + \lambda_1 i)}} \\ &= -H(Q_X) + \sum_i (\lambda_0 + \lambda_1 \, i) \, Q_X(i) \\ &= -H(Q_X) + \lambda_0 + \lambda_1 \alpha, \end{aligned}$$

where the last equality holds because the expected value under Q_X is α. Therefore, [30%]

$$H(Q_X) \leq \lambda_0 + \lambda_1 \alpha \tag{2}$$

Moreover, the entropy of P_X^* is

$$H(P_X^*) = \sum_i 2^{-(\lambda_0 + \lambda_1 i)} \, (\lambda_0 + \lambda_1 i) = \lambda_0 + \lambda_1 \alpha. \tag{3}$$

Combining (2) and (3) we see that $H(P_X^*) \geq H(Q_X)$, with equality if and only if $Q_X = P_X^*$.

(c) From part (i), the max-entropy pmf P_X^* over $\{1, 2\}$ is of the form

$$P_X^*(1) = 2^{-(\lambda_0 + \lambda_1)} = p, \qquad P_X^*(2) = 2^{-(\lambda_0 + 2\lambda_1)} = pq, \quad P_X^*(2) = 2^{-(\lambda_0 + 3\lambda_1)} = pq^2$$

where we have denoted $p = 2^{-(\lambda_0 + \lambda_1)}$ and $q = 2^{-\lambda_1}$. For it be a valid pmf, we need

$$p + pq + pq^2 = 1 \quad \Rightarrow \quad p = \frac{1}{1 + q + q^2}. \tag{4}$$

The expected value is

$$E[X] = 1p \, + \, 2pq + \, + 3pq^2 \; = p(1 + 2q + 3q^2) = \frac{(1 + 2q + 3q^2)}{(1 + q + q^2)} = 1.5 \tag{5}$$

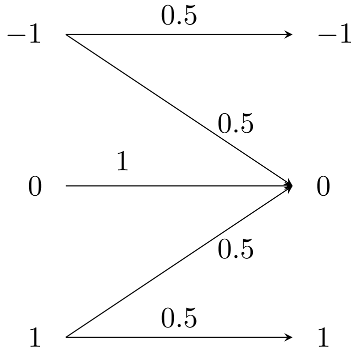
Therefore q is the solution of the quadratic $1.5q^2 + 0.5q - 0.5 = 0$. This gives $q = 0.4343$, from which $p = \frac{1}{1+q+q^2} = 0.6162$. Therefore the max-entropy pmf is [25%]

$$P_X^*(1) = 0.6162, \quad P_X^*(2) = 0.2676, \quad P_X^*(3) = 0.1162.$$

Assessor's comment: In part (a).(i), most stopped with Eq. (1) and did not simplify the condition to explicitly specify which sequences are contained in the typical set. This led to inaccurate answers for (a).(ii), based on bounds on the size of the typical set rather than the exact number of sequences in it. Part (b).(i) was well answered in general, but most did not solve the simultaneous equations in (b).(ii), perhaps due to lack of time.

Question 3

(a) i) With the given $Z-Y$ mapping, the transition probability matrix is [20%]



Transition probability matrix:

		Y		
	$P_{Y\mid X}$	-1	0	1
X	-1	0.5	0.5	0
	0	0	1	0
	1	0	0.5	0.5

ii) Due to the symmetry between inputs -1 and 1, the optimal output distribution will have the form [30%]

$P_X(-1) = P_X(1) = p$ and $P_X(0) = 1-2p$. The corresponding output distribution is given by $P_Y(-1) = P_Y(1) = \frac{p}{2}$ and $P_Y(0) = (1-p)$. Then the mutual information is

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y \mid X) \\
&= H(\{p/2,\, p/2,\, 1-p\}) - p\underbrace{H(Y \mid X=-1)}_{1} - (1-2p)\underbrace{H(Y \mid X=0)}_{0} - p\underbrace{H(Y \mid X=1)}_{1} \\
&= 2 \cdot \frac{p}{2}\log_2\frac{2}{p} + (1-p)\log_2\frac{1}{(1-p)} \;-\; 2\cdot p \cdot 1 \\
&= H_2(p) - p.
\end{aligned}$$

We want to maximise $f(p) = H_2(p) - p$ over $p \in (0, 1/2)$. Setting the derivative to zero:

$$f'(p) = \log_2\frac{1-p}{p} \;-\; 1 = 0 \quad \Rightarrow \quad \text{Optimal value of } p = p^* = \frac{1}{3}$$

Capacity $= H_2(1/3) - 1/3 = 0.585$ bits. Capacity achieving distribution $P_X(-1) = P_X(0) = P_X(1) = \frac{1}{3}$.

(b) i) Using the law of total probability, the density of $Z = X + N$ can be written as [20%]

$$\begin{aligned}
f(z) &= pf(z \mid X=-1) \;+\; (1-2p)f(z \mid X=0) \;+\; pf(z \mid X=1) \\
&= \begin{cases} p/2, & \text{for } -2 \le z < 1, \\ (1-p)/2, & \text{for } -1 \le z \le 1, \\ p/2, & \text{for } 1 < z \le 2. \end{cases}
\end{aligned}$$

ii) The mutual information with an input distribution of the form in part (i) is [30%]

$$\begin{aligned}
I(X;Z) &= h(Z) - h(Z \mid X) \\
&= h(Z) - ph(Z \mid X=-1) - (1-p)h(Z \mid X=0) - ph(Z \mid X=1) \\
&= h(Z) - p\log_2 2 - (1-p)\log_2 2 - p\log_2 2 = h(Z) - 1.
\end{aligned}$$

In the last line, we have used the fact that the differential entropy of a uniform random variable on $[A, B]$ is $\log_2[B-A]$. Z can take values in the interval $[-2, 2]$, and from the hint, $h(Z)$ is maximised when Z is uniform in $[-2, 2]$. (In this case $h(Z) = \log_2 4$.)

From the density $f(z)$ computed in part (i), we see that the density is Unif$[-2, 2]$ if $p/2 = (1-p)/2$, i.e., for $p = 1/2$.

Therefore, a capacity achieving input distribution is $P_X(-1) = P_X(1) = \frac{1}{2}$ and $P_X(0) = 0$. The capacity is

$$\mathcal{C} = \log_2 4 - 1 = 1 \text{ bit.}$$

Assessor's comment: Parts (a).(i) and (b).(i) were done well in general, but relatively few got the correct channel capacity in (a).(ii), often due to not using the symmetry between the inputs -1 and 1.

Question 4

(a) Code length $n = 6$, $(n-k) = 4$, therefore dimension $n = 2$. Rate $= 2/6 = 1/3$. [10%]

(b) We perform elementary row operations to get a 4×4 identity matrix on the right: [20%]

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{R_4 \leftarrow R_4 + R_3} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{R_1 \leftarrow R_1 + R_2,\, R_3 \leftarrow R_3 + R_2} \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Finally push rows 3,4 to the top to get:

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

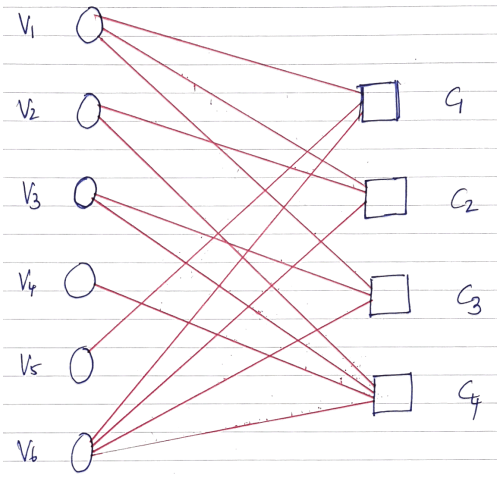
(c) If $\mathbf{H}_{sys} = [P^T \mid I_{n-k}]$, then $\mathbf{G}_{sys} = [I_k \mid P]$. Hence [10%]

$$\mathbf{G}_{sys} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(d) The code has 4 codewords. These can be found from $\mathbf{G}_{sys}$. The codewords corresponding using to the information sequences $[0,0]$, $[0,1]$, $[1,0]$ and $[1,1]$ are $[0,0,0,0,0,0]$, $[1,0,0,1,0,1]$, $[0,1,1,1,1,1]$, and $[1,1,1,0,1,0]$, respectively. [10%]

Therefore, for the received sequence $[1,1,1,1,1,1]$ the minimum-distance codeword is $[0,1,1,1,1,1]$.

(e) The factor graph is shown below: [15%]



(f) i) The channel evidence in LLR form for the AWGN channel is (from databook) $L(y_j) = \frac{2}{\sigma^2} y_j$. For the six variable nodes, these are: $L(\underline{y}) = [-1, 1, 2.4, 1.2, 0.4, 1.6]$. [5%]

The message sent by variable node 1 to checks $1, 5$, and 6 in the first iteration is $L(y_1) = -1$.

ii) The fourth code bit (denoted v_4) is connected only to the fourth check node (denoted c_4).

The messages received by c_4 from v_2, v_3, v_6 in the first iteration are $L(y_2) = 1, L(y_3) = 2.4, L(y_6) = 1.6$, respectively. [30%]

Therefore the message $L_{c_4 \to v_4}$ in the first iteration is:

$$L_{c_4 \to v_4} = 2 \tanh^{-1} \left(\tanh(0.5) \tanh(1.2) \tanh(0.8) \right) = 0.5233.$$

Therefore the final LLR for code bit 4 after one round of message passing is $\hat{L}_4 = L(y_4) + L_{c_4 \to v_4} = 1.7233$. Since this is positive, the fourth code bit is decoded as a 0.

We follow the same procedure for the fifth code bit (v_5) which is connected only to the check c_1.

The messages received by c_1 from v_1, v_6 in the first iteration are $L(y_1) = -1, L(y_6) = 1.6$, respectively. Therefore the message $L_{c_1 \to v_5}$ in the first iteration is:

$$L_{c_1 \to v_5} = 2 \tanh^{-1} \left(\tanh(-0.5) \tanh(0.8) \right) = -0.6342.$$

Therefore the final LLR for code bit 5 after one round of message passing is $\hat{L}_5 = L(y_5) + L_{c_4 \to v_4} = -0.2342$. Since this is negative, the fifth code bit is decoded as a 1.

Assessor's comment: The most popular question in the paper and well-answered by most. Many students did not use the formula from the data book for the channel log-likelihood ratios in part (f), and computed these from scratch.