3F7 Information Theory and Coding
Engineering Tripos 2024/25 – Solutions

Question 1

(a) (i) The probability density of y, denoted $f(y)$, can be computed as

$$f(y) = \underbrace{P_X(A)}_{2/3} f(y \mid X = A) + \underbrace{P_X(B)}_{1/3} f(y \mid X = B).$$

We have $f(y \mid X = A) = \frac{1}{2}$ for $y \in [-1, 1]$, and $f(y \mid X = B) = 1$ for $y \in [-\frac{1}{2}, \frac{1}{2}]$. Using this we have [15%]

$$f(y) = \begin{cases} \frac{1}{3}, & -1 \le y < \frac{1}{2}, \\ \frac{2}{3}, & -\frac{1}{2} \le y \le \frac{1}{2}, \\ \frac{1}{3}, & \frac{1}{2} < y \le 1. \end{cases}$$

(ii) The mutual information $I(X;Y) = h(Y) - h(Y \mid X)$. Using the expression above for $f(y)$, [15%]

$$\begin{aligned} h(Y) &= \int_{-1}^{1} f(y) \log \frac{1}{f(y)} dy \\ &= \int_{-1}^{-\frac{1}{2}} \frac{1}{3} \log 3 \, dy \; + \; \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{2}{3} \log \frac{3}{2} \, dy \; + \; \int_{\frac{1}{2}}^{1} \frac{1}{3} \log 3 \, dy = \log 3 - \frac{2}{3}. \end{aligned}$$

The conditional entropy is

$$h(Y \mid X) = \frac{2}{3} h(Y \mid X = A) + \frac{1}{3} h(Y \mid X = B) = \frac{2}{3} \log 2 + \frac{1}{3} \log 1 = \frac{2}{3}.$$

Therefore the mutual information is $I(X;Y) = \log 3 - \frac{4}{3} = 0.2516$ bits.

(b) Using the hint, the cumulative distribution function $F_Y(y) = F_X(y/a)$. Differentiating with respect to y, the pdfs are related as $f_Y(y) = \frac{1}{a} f_X(y/a)$. Therefore [20%]

$$\begin{aligned} h(Y) = \int f_Y(y) \log \frac{1}{f_Y(y)} dy &= \int \frac{1}{a} f_X(y/a) \log \frac{a}{f_X(y/a)} dy \\ &= \int f_X(u) \log \frac{a}{f_X(u)} du \qquad \text{(substituting } u = y/a\text{)} \\ &= \int f_X(u) (\log a) \, du \; + \; h(X) = h(X) + \log a. \quad \square \end{aligned}$$

(c) (i) Using the chain rule we have $H(X,Y,Z) = H(X,Y) + H(Z \mid X,Y)$ and $H(X,Z) = H(X) + H(Z \mid X)$. Since conditioning can only reduce entropy, we have $H(Z \mid X,Y) \le H(Z \mid X)$. [15%]

Therefore, $H(X,Y,Z) - H(X,Y) \; \le \; H(X,Z) - H(X)$.

(ii) Here the two sides are equal. Two see this, we expand the LHS as: [15%]

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

The RHS is the same expression since

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

(d) We expand $H(X,Y,Z)$ in two different ways: [20%]

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

Adding the two equations, we obtain

$$2H(X,Y,Z) = H(X,Y) + H(Y,Z) + H(Z \mid X,Y) + H(X \mid Y,Z).$$

Therefore to get the result, we need to show that $H(Z \mid X,Y) + H(X \mid Y,Z) \leq H(Z,X)$. This follows from chain rule and the fact that conditioning can only reduce the entropy:

$$H(Z,X) = \underbrace{H(Z)}_{\geq H(Z|X,Y)} + \underbrace{H(X \mid Z)}_{\geq H(X|Y,Z)} \geq H(Z \mid X,Y) + H(X \mid Y,Z).$$

Assessor's comment: Many students made struggled with part (b) as they did not correctly express the density of Y in terms of that of X.

Question 2

(a) (i) The entropy $H(X)$ is a lower bound on minimum expected code length/source symbol. [10%]

$$H(X) = -(0.99)\log_2(0.99) - (0.01)\log_2(0.01) = 0.0808.$$

(ii) The number of sequences with at least 97 a symbols, denoted by N_0 is [10%]

$$N_0 = \binom{100}{97} + \binom{100}{98} + \binom{100}{99} + \binom{100}{100} = 166,751.$$

To assign an equal length codeword to each of these sequences, we need $\lceil \log_2 166,751 \rceil = \lceil 17.34 \rceil = 18$ bits.

(iii) The probability that the source sequence has at least 97 a symbols, denoted by q, is [20%]

$$q = \binom{100}{97}(0.99)^{97}(0.01)^3 + \binom{100}{98}(0.99)^{98}(0.01)^2 + \binom{100}{99}(0.99)^{99}(0.01)^+ \binom{100}{100}(0.99)^{100}$$
$$= 0.9816.$$

The expected code length is $L_n = q(18) + (1-q)100 = 19.51$, or expected code length per source symbol $= 0.1951$.

(iv) For the expected code length to approach the entropy, we need to code over much large source sequences, i.e., take $n \gg 100$. Then, in step ii), taking a small $\epsilon > 0$, we assign an equal length codeword to sequences for which the number of ones is in the range $[(0.99 - \epsilon)n, (0.99 + \epsilon)n]$. This set has not more than $2^{n(H(X)+\epsilon)}$ sequences and for large n, this set has very high probability (at least $(1-\epsilon)$). The remaining sequences just get length-n codewords. The expected code length is then: [15%]

$$\bar{L}_n < n(H(X) + \epsilon) + \epsilon n \qquad \frac{\bar{L}_n}{n} < H(X) + \epsilon' \text{ bits/symbol.}$$

(One could also using arithmetic coding to attain an expected code length close to the entropy.)

(b) (i) Let A be the most probably symbol, with probability greater than $\frac{2}{5} = 0.4$, and assume towards contradiction that its codeword has length at least 2. Then at some stage A must have been merged with another 'super-symbol', say S_0. (Super-symbol is either a symbol or what is obtained after one or more merges.) Moreover, there must also be a third super-symbol, say S_1, which has probability greater than 0.4, as otherwise, A and S_0 wouldn't be merged. Furthermore, S_1 must have been obtained by one or more merges, as it has probability greater than A. [25%]

These lead to two observations: 1) S_0 has probability less than 0.2 (since A and S_1 each have probability at least 0.4); and 2) among the two components that are merged to form S_1, call them S_{10} and S_{11}, the larger of them (say S_{11}) has at least half the probability of S_1, i.e., the probability of S_{11} is greater than 0.2. And the smaller of them S_{10} must have probability at most 0.3 (as the total probability of everything other than A is at most 0.6). 1) and 2) together lead to a contradiction, as they imply that Huffman procedure should have merged S_0 and S_{10} rather than S_{11} and S_{10}. Therefore the initial assumption of A having a codeword obtained via a merge is false.

(ii) We again prove by contradiction. Assume that the most probable symbol A with probability $p_1 < \frac{1}{3}$ has a codeword of length 1. Assume that at the last stage, it is merged with super-symbol S, whose total probability is $1 - p_1 > \frac{2}{3}$. S is obtained by merging two components,

call them S_0 and S_1. The component with larger probability, say S_1, has at least half the probability of S, i.e, $P(S_1) > \frac{1}{3}$. This leads to a contradiction since the Huffman procedure should have merged S_0 and A rather than S_0 and S_1 (since $P(A) < \frac{1}{3}$ and $P(S_1) > \frac{1}{3}$). Therefore the initial assumption of A having a codeword of length 1 is false. [20%]

Assessor's comment: Part (a) is similar to a question in Examples Paper 1, but there were a surprising number of long-winded, erroneous answers. Part (b), on proving that Huffman codeword lengths satisfy certain properties, was probably the hardest question in the exam. Many gave reasonable answers that went in the right direction, and pleasingly, some gave a perfect proof for part (b)(ii).

Question 3

(a) We have $I(X;Y) = H(Y) - H(Y \mid X) \leq H(Y) \leq 1$, where the last inequality follows Y can take only two values. Therefore the capacity $\mathcal{C} \leq \max_{P_X} I(X.;Y) \leq 1$. By taking the input distribution $P_X(0) = P_X(1) = \frac{1}{2}$ and $P_X(2) = 0$, we have $H(Y \mid X) = 0$ and $H(Y) = 1$ (since the induced distribution on Y is $P_Y(0) = P_Y(1) = \frac{1}{2}$. Therefore, with this input distribution, $I(X;Y) = 1$,] which implies that the capacity is 1 bit. [15%]

(b) (i) Let the input distribution be $P(X = 1) = \alpha$, $P(X = 0) = 1 - \alpha$. Considering a Z-channel with $P(Y = 0 \mid X = 1) = p$ (in the question $p = 1/3$), the output distribution is

$$P(Y = 1) = P(X = 0)\underbrace{P(Y = 1|X = 0)}_{0} + P(X = 1)P(Y = 1|X = 1) = \alpha(1 - p). \tag{1}$$
$$P(Y = 0) = (1 - \alpha) + \alpha p.$$

We then have [25%]

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - [P(X = 0)H(Y|X = 0) + P(X = 1)H(Y|X = 1)] \\ &= H(Y) - (1 - \alpha)0 - \alpha H_2(p) \\ &= H_2(\alpha(1 - p)) - \alpha H_2(p). \end{aligned} \tag{2}$$

To find the capacity, we have to maximize $I(X;Y)$ with respect to $\alpha \in (0, 1)$. Differentiating the above w.r.t α and setting to zero, we obtain

$$\frac{\mathrm{d}}{\alpha} I(X;Y) = (1 - p) \log_2 \frac{(1 - \alpha) + \alpha p}{\alpha(1 - p)} \; - \; H_2(p) = 0.$$

Solving this with $p = 1/3$ gives the maximizer $\alpha^* = 0.4169$. The capacity is $\mathcal{C} = H_2(\alpha^* 2/3) - \alpha * H_2(1/3) = 0.4697$ bits/channel use.

(ii) For the cascade with m independent Z-channels, we have $P(Y = 0 \mid X = 0) = 1$, and $P(Y = 1 \mid X = 1) = (2/3)^m$ and $P(Y = 1 \mid X = 1) = 1 - (2/3)^m$. [15%]
As $m \to \infty$, $P(Y = 1 \mid X = 1) \to 1$, therefore the capacity tends to 0 (as both inputs give the same output).

(c) (i) We have

$$\begin{aligned} I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(X + Z \mid X) &= H(Y) - H(Z) \\ &= H(Y) - \log 3 \end{aligned} \tag{3}$$

where the last equality holds because N is uniformly distributed among $\{0, 1, 2\}$. The maximum value of $H(Y)$ is $\log 24$, which is attained when Y is uniformly distributed in $\{0, 1, 2, \ldots, 23\}$. This can be achieved by taking the input X to be uniform over $\{0, 1, 2, \ldots, 23\}$. Indeed, with $P_X(k) = \frac{1}{24}$ for $k \in \{0, 1, \ldots, 23\}$, we have

$$P_Y(y) = \frac{1}{3}P_X(y) + \frac{1}{3}P_X(y - 1) + \frac{1}{3}P_X(y - 2) = \frac{1}{24}, \quad y \in \{0, \ldots, 23\}.$$

Here $(y-1)$ and $(y-2)$ are computed modulo 4. The capacity is therefore $\mathcal{C} = \log 24 - \log 3 = \log(8) = 3$ bits per channel use. [20%]
Alternative input distribution: The same output distribution on Y can also be obtained by taking X uniform over the eight symbols $\{0, 3, 6, 9 \ldots, 21\}$ (similar to the noisy keyboard channel in Handout 7).

(ii) To transmit three bits per channel use, the encoder maps each of the 8 three-bit patterns to a unique number in $\{0, 3, 6, 9 \ldots, 21\}$, e.g., $000 \to 0$, $001 \to 3$, $\ldots$,$111 \to 7$. This set of inputs is *non-confusable*, i.e., each input symbol can be exactly recovered from the output symbol. For example, an output $0, 1$, or 2 implies that the input symbol is 0, an output 3/4/5 implies that the input symbol is 1. This gives an error-free scheme to transmit at 3 bits per channel use. [25%]

Assessor's comment: Many spent more time and effort than needed on part (a), giving wrong answers by incorrectly assuming symmetry in the input distribution.

Question 4

(a) Consider any two codewords $\mathbf{c}_1 = \mathbf{aG}$ and $\mathbf{c}_2 = \mathbf{bG}$, where $\mathbf{a} = (a_1, \ldots, a_k)$ and $\mathbf{b} = (b_1, \ldots, b_k)$ are the corresponding binary vectors of information bits. Letting the rows of the binary $n \times k$ generator matrix $\mathbf{G}$ be $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_k$, we have

$$\mathbf{c}_1 = a_1\mathbf{g}_1 + a_2\mathbf{g}_2 + \ldots + a_k\mathbf{g}_k, \qquad \mathbf{c}_2 = b_1\mathbf{g}_1 + b_2\mathbf{g}_2 + \ldots + b_k\mathbf{g}_k,$$

The sum is

$$\mathbf{c}_1 + \mathbf{c}_2 = (a_1 + b_1)\mathbf{g}_1 + (a_2 + b_2)\mathbf{g}_2 + \ldots + (a_k + b_k)\mathbf{g}_k.$$

Since the addition is modulo 2, we observe that $\mathbf{c}_1 + \mathbf{c}_2$ is a codeword corresponding to the vector of information bits $((a_1 + b_1), (a_2 + b_2), \ldots, (a_k + b_k))$. Therefore the sum of any two codewords is a codeword. [20%]

Scalar multiplication: For any codeword $\mathbf{c}$, we have $1 \cdot \mathbf{c} = \mathbf{c}$, and $0 \cdot \mathbf{c} = \mathbf{0}$. The all zeros vector of length n is a codeword, corresponding to the all-zeros information sequence, which shows that multiplying a codeword by a scalar in $\{0, 1\}$ gives a codeword.

(b) (i) $n = 5$, $n - k = 3$, therefore the dimension $k = 2$ and rate $k/n = 2/5$. [10%]

(ii) Since $f(y \mid x = -1) = f(-y \mid x = 1)$ [15%]

$$f(y \mid x = 1) = \begin{cases} \frac{1}{3}e^{-y/2}, & y \geq 0, \\ \frac{1}{3}e^{y}, & y < 0, \end{cases} \qquad f(y \mid x = -1) = \begin{cases} \frac{1}{3}e^{-y}, & y > 0, \\ \frac{1}{3}e^{y/2}, & y \leq 0. \end{cases}$$

Therefore, the log-likelihood ratio (LLR) is

$$L(y) := \ln \frac{f(y \mid x = 1)}{f(y \mid x = -1)} = \frac{y}{2}, \quad y \in \mathbb{R}.$$

(iii) The message transmitted by variable node j in the first iteration is the LLR $L(y_j) = y_j/2$, for $j = 1, \ldots, 5$. For the given received sequence, these LLRs are

$$[L(y_1), \ldots, L(y_5)] \; = \; [1.05, \, -0.05, \, 0.3, \, 0.2, \, -0.75].$$

The fourth code bit is connected to the first two second check nodes. Therefore its LLR after one complete round of message passing is:

$$L_4 = L(y_4) + L_{c_1 \to v_4} + L_{c_2 \to v_4} = 0.2 + + L_{c_1 \to v_4} + L_{c_2 \to v_4},$$

where $L_{c_1 \to v_4}, L_{c_2 \to v_4}$ are the messages sent from check nodes 1 and 2, respectively, to v_4. [25%] The message $L_{c_2 \to v_4}$ is determined by the initial LLRs received from v_1 and v_2 (the v-nodes other than v_4 connected to c_2):

$$L_{c_2 \to v_4} = 2\tanh^{-1}\left[\tanh(L(y_1)/2)\tanh(L(y_2)/2)\right] = 2\tanh^{-1}\left[\tanh(0.525)\tanh(-0.025)\right] = -0.0241.$$

Similarly, $L_{c_1 \to v_4} = 2\tanh^{-1}\left[\tanh(L(y_3)/2)\right] = L(y_3) = 0.3$. Using these, the LLR for the fourth code bit after one complete round is $L_4 = 0.2 - 0.0241 + 0.3 = 0.4759$. Since it is positive the bit will be decoded to a 0.

(iv) A systematic parity matrix (using rows operations $R_3 \to R_1 + R_2 + R_3$ and $R_1 \to R_2 + R_1$) and a systematic generator matrix can be obtained as: [10%]

$$\mathbf{H}_{sys} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{G}_{sys} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{4}$$

Using this, the four codewords can be determined as:

$$[0, 0, 0, 0, 0], \; [1, 0, 1, 1, 0], \; [0, 1, 1, 1, 1], \; [1, 1, 0, 0, 1].$$

(v) The ratio of the posterior probabilities for the fourth code bit given $\underline{y}$ can be expressed as:

$$\frac{P(c_4 = 0 \mid \underline{y})}{P(c_4 = 1 \mid \underline{y})} = \frac{\sum_{\underline{c}:c_4=0} P(\underline{c} \mid \underline{y})}{\sum_{\underline{c}:c_4=1} P(\underline{c} \mid \underline{y})} = \frac{\sum_{\underline{x}:x_4=1} f(\underline{y} \mid \underline{x})}{\sum_{\underline{x}:x_4=-1} f(\underline{y} \mid \underline{x})}, \tag{5}$$

where the last equality follows because $P(\underline{c})$ is the same for each codeword, and that $f(y_j \mid c_j = 0) = f(y_j \mid x_j = 1)$ and $f(y_j \mid c_j = 1) = f(y_j \mid x_j = -1)$. Using the conditional density of the channel in part (ii), we have: [20%]

$$f(\underline{y} \mid \underline{x} = [1,1,1,1,1]) = (1/3)^5 \exp(-1.05 - 0.1 - 0.3 - 0.2 - 1.5) = c \cdot 0.0429,$$
$$f(\underline{y} \mid \underline{x} = [-1,1,-1,-1,1]) = c \cdot \exp(-2.1 - 0.1 - 0.6 - 0.4 - 1.5) = c \cdot 0.0091,$$
$$f(\underline{y} \mid \underline{x} = [1,-1,-1,-1,-1]) = c \cdot \exp(-1.05 - 0.05 - 0.6 - 0.4 - 0.75) = c \cdot 0.0578,$$
$$f(\underline{y} \mid \underline{c} = [-1,-1,1,1,-1]) = c \cdot \exp(-2.1 - 0.05 - 0.3 - 0.2 - 0.75) = c \cdot 0.0334$$

(Here $c = (1/3)^5$.) Using this in (5) we obtain

$$\frac{P(c_4 = 0 \mid \underline{y})}{P(c_4 = 1 \mid \underline{y})} = \frac{0.0429 + 0.0334}{0.0091 + 0.0578} = 1.1405.$$

(Optional): Taking logs, we see that $\ln \frac{P(c_2=0|\underline{y})}{P(c_2=1|\underline{y})} = 0.1315$, which is smaller than the LLR obtained after one round of message passing.

Assessor's comment: Many students got the wrong expression for the log-likelihood ratio in part (b)(ii), due to not being careful with how the expression for the likelihood changes with the sign of y.