3F7 Information Theory and Coding
Engineering Tripos 2023/24 – Solutions

Question 1

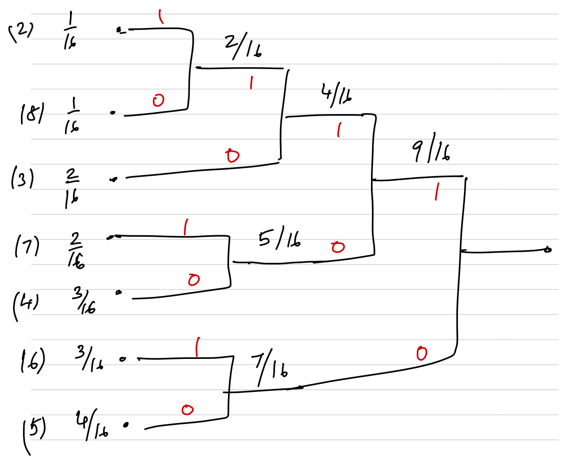
(a) Z can take values in $\{2, \ldots, 8\}$. The probability mass function is [20%]

$$P(Z=2) = P(X=1, Y=1) = \frac{1}{4}\frac{1}{4} = \frac{1}{16},$$
$$P(Z=3) = P(X=1, Y=2) + P(X=2, Y=1) = \frac{1}{4}\frac{1}{4} + \frac{1}{4}\frac{1}{4} = \frac{2}{16},$$
$$P(Z=4) = P(X=1, Y=3) + P(X=3, Y=1) + P(X=2, Y=2) = \frac{3}{16},$$
$$P(Z=5) = P(X=2, Y=3) + P(X=3, Y=2) + P(X=1, Y=4) + P(X=4, Y=1) = \frac{4}{16},$$
$$P(Z=6) = P(X=4, Y=2) + P(X=2, Y=4) + P(X=3, Y=3) = \frac{3}{16},$$
$$P(Z=7) = P(X=4, Y=3) + P(X=3, Y=4) = \frac{2}{16}, \quad P(Z=8) = P(X=4, Y=4) = \frac{1}{16}.$$

(b) If $Z = k$, then $(k-1)$ questions are needed, for $k \in \{1, \ldots, 7\}$. If $Z = 8$, then it detected after 6 questions (Is $Z = 7$?). So the expected number of questions is: [20%]

$$\bar{L} = 1 \cdot \frac{1}{16} + 2 \cdot \frac{2}{16} + 3 \cdot \frac{3}{16} + 4 \cdot \frac{4}{16} + 5 \cdot \frac{3}{16} + 6 \cdot \frac{2}{16} + 6 \cdot \frac{1}{16} = \frac{63}{16} = 3.94.$$

(c) The optimal strategy is given by the Huffman code, constructed as shown below. [25%]



$$5 \to 00, \quad 6 \to 01, \quad 4 \to 100, \quad 7 \to 101, \quad 3 \to 110, \quad 8 \to 1110, \quad 2 \to 1111.$$

The expected number of questions with this strategy is

$$L^* = 2\left(\frac{3}{16} + \frac{4}{16}\right) + 3\left(\frac{3}{16} + \frac{2}{16} + \frac{2}{16}\right) + 4\left(\frac{1}{16} + \frac{1}{16}\right) = \frac{43}{16} = 2.69.$$

Starting from the root, the split at each level specifies the question. E.g., the first question is "Is $Z \in \{5,6\}$?"; if no, then the next question is: $Z \in \{4,7\}$?", and so on.

(There are other valid Huffman trees, which may give strategies with a different sequence of questions, but they will all have the same expected number of questions.)

(d) The conditional joint entropy can be expressed as

$$\begin{aligned} H(X,Y \mid Z) &= H(X \mid Z) + H(Y \mid X,Z) \\ &= H(X \mid Z) \quad \text{since } Y \text{ is a function of } (X,Z) \\ &= \sum_{k=2}^{8} P(Z=8) \cdot H(X \mid Z=k) \end{aligned}$$

When $Z = 2$, we have $X = 1, Y = 1$, therefore $H(X \mid Z = 2) = 0$. Similarly, $H(X \mid Z = 8) = 0$.

When $Z = 3$, we have one of two choices: $(X = 1, Y = 2)$, $(X = 2, Y = 1)$, both of which have equal probability. Therefore, $H(X \mid Z = 3) = 1$. Similarly, $H(X \mid Z = 7) = 1$. [30%]

When $Z = 4$, we have one of three choices: $(X = 2, Y = 2)$, $(X = 1, Y = 3)$ and $(X = 3, Y = 1)$, all of which have equal probability. Therefore, $H(X \mid Z = 4) = \log_2 3$. Similarly, $H(X \mid Z = 6) = \log_2 3$. Finally, for $Z = 5$, we have four choices with equal probability. Therefore, $H(X \mid Z = 5) = \log_2 4 = 2$.

Putting everything together, we obtain:

$$H(X,Y \mid Z) = \frac{2}{16}(1+1) \; + \; \frac{3}{16}(\log_2 3 + \log_2 3) \; + \; \frac{4}{16}(2) = 1.344 \text{ bits}$$

Alternative approach: Using the chain rule, $H(X,Y,Z)$ can be written in two ways:

$$H(X,Y,Z) = H(Z) + H(X,Y \mid Z) = H(X,Y) \; + \; \underbrace{H(Z \mid X,Y)}_{0}$$

Using the pmf $P(Z)$ above, we find that $H(Z) = 2.656$. We therefore have

$$H(X,Y \mid Z) = H(X,Y) - H(Z) = H(X) + H(Y) - H(Z) = 2 + 2 - 2.656 = 1.344.$$

Question 2

(a) i) The codeword lengths for the Shannon-Fano code are: [25%]

$$\ell_1 = \lceil \log_2(2^{k_1}) \rceil = k_1,\ \ell_2 = \lceil \log_2(2^{k_2}) \rceil = k_2,\ \ldots,\ \ell_M = \lceil \log_2(2^{k_M}) \rceil = k_M.$$

The expected codelength is $\bar{L} = \sum_{i=1}^{M} \ell_i P(i) = \sum_{i=1}^{M} k_i 2^{-k_i}$.

The entropy $H(X) = \sum_{i=1}^{M} P(i) \log_2(1/P(i)) = \sum_{i=1}^{M} k_i 2^{-k_i}$. Therefore, $\bar{L} = H(X)$, i.e., the Shanon-Fano code has the smallest possible expected code length.

ii) The codeword lengths of the new code are: [15%]

$$\ell_1' = \lceil \log_2(2^{c_1}) \rceil = c_1,\ \ell_2' = \lceil \log_2(2^{c_2}) \rceil = c_2,\ \ldots,\ \ell_M' = \lceil \log_2(2^{c_M}) \rceil = c_M.$$

The expected codelength is $\bar{L}' = \sum_{i=1}^{M} P(i)\ell_i' = \sum_{i=1}^{M} c_i 2^{-k_i}$.

iii) The difference in the expected codelengths is

$$\bar{L}' - \bar{L} = \sum_{i=1}^{M} (c_i - \ell_i) 2^{-k_i}.$$

This is also equal to the relative entropy $D(P\|Q)$ since

$$D(P\|Q) = \sum_{i=1}^{M} P(i) \log_2 \frac{P(i)}{Q(i)} = \sum_{i=1}^{M} 2^{-k_i} \log_2 \frac{2^{-k_i}}{2^{-c_i}} = \sum_{i=1}^{M} (c_i - \ell_i) 2^{-k_i}.$$

Since $D(P\|Q) \geq 0$ with equality if and only if $P = Q$, it follows that $\bar{L}' - \bar{L} \geq 0$, with equality if and only if $P = Q$. [20%]

(b) i) The probability mass function of X_1 is

$$P(X_1 = r) = \frac{r}{r+w+b}, \quad P(X_1 = w) = \frac{w}{r+w+b}, \quad P(X_1 = b) = \frac{b}{r+w+b}.$$

Therefore the entropy is [10%]

$$H(X_1) = \frac{r}{r+w+b} \log_2 \frac{r+w+b}{r} + \frac{w}{r+w+b} \log_2 \frac{r+w+b}{w} + \frac{b}{r+w+b} \log_2 \frac{r+w+b}{b}.$$

ii) We have

$$H(X_1, \ldots, X_K) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_2, X_1) + \ldots + H(X_K \mid X_{K-1}, \ldots, X_1). \quad (1)$$

When the balls are drawn with replacement, $X_1, X_2, \ldots, X_K$ are mutually independent and have the same distribution. Therefore, [10%]

$H(X_2 \mid X_1) = H(X_2)$, $H(X_3 \mid X_2, X_1) = H(X_3)$, $\ldots$, $H(X_K \mid X_{K-1}, \ldots, X_1) = H(X_K)$, and hence $H(X_1, \ldots, X_K) = KH(X_1)$, where $H(X_1)$ was computed in part (i).

iii) We have

$$H(Y_1, \ldots, Y_K) = H(Y_1) + H(Y_2 \mid Y_1) + H(Y_3 \mid Y_2, Y_1) + \ldots + H(Y_K \mid Y_{K-1}, \ldots, Y_1). \quad (2)$$

The pmf for Y_1 is the same as that of X_1, hence $H(X_1) = H(Y_1)$. Without replacement Y_2 depends on Y_1, Y_3 depends on (Y_1, Y_2), and so on. Since conditioning can only reduce entropy, we have

$$H(Y_2 \mid Y_1) \leq H(Y_2) < H(Y_1) = H(X_1),$$

where $H(Y_2) < H(Y_1)$ will hold because Y_2 has one fewer choice than Y_1 (you don't need to prove this rigorously). Considering the conditional entropies one by one, we obtain [20%]

$H(Y_3 \mid Y_2, Y_1) < H(X_1)$, $\ldots H(Y_K \mid Y_{K-1}, \ldots, Y_1) < H(X_1)$. Therefore $H(X_1, \ldots, X_K) > H(Y_1, \ldots, Y_K)$.

Question 3

(a) Let the density of X be f, and let u be the density of the uniform random variable on $[A, B]$, i.e., $u(x) = \frac{1}{(B-A)}$ for $x \in [A, B]$. Since the relative entropy $D(f\|u) \geq 0$, we have: [20%]

$$\begin{aligned}
0 \leq D(f\|u) &= \int_A^B f(x) \log \frac{f(x)}{u(x)} dx \\
&= \int_A^B f(x) \log \left[f(x)(B-A) \right] \, dx \\
&= \int_A^B f(x) \log(B-A) dx \; + \; \int_A^B f(x) \log f(x) dx \\
&= \log(B-A) - h(X) \qquad \text{using} \int f(x) dx = 1, \text{ and } h(X) = -\int f(x) \log(f(x)) dx
\end{aligned}$$

Therefore $h(X) \leq \log(B - A)$, with equality if and only if $f = u$, since $D(f\|u) = 0$ if and only if $f = u$.

(b) The mutual information $I(V; Y)$ can be written as [25%]

$$\begin{aligned}
I(V;Y) &= h(Y) - h(Y \mid V) \\
&= h(Y) - h(Z) \quad [\text{ since } Y = Z + V \text{ with } Z \text{ independent of } V] \\
&= h(V + Z) - \log_2 2 \quad [\text{ since } Z \text{ is uniform in the interval } [-1, \, 1]] \\
&\leq \log_2(4) - 1 = 1
\end{aligned}$$

Since $V \in \{1 - 1/\}$ and Z is uniform in $[-1, \, 1]$, we have $-2 \leq Y \leq 2$, and using part (a), $h(Y) \leq \log_2(4)$, with equality if Y is uniform in the interval $[-2, 2]$. If we choose the input distribution $P(V = -1) = P(V = 1) = 0.5$, then Y is indeed uniform in $[-2, 2]$. To see this, the probability density function $f_Y(y)$ is

$$\begin{aligned}
f_Y(y) &= P(V=1) f_{Y|Z}(y \mid Z = 1) \; + \; P(V=-1) f_{Y|Z}(y \mid Z = -1) \\
&= \frac{1}{2} f_Z(y-1) + \frac{1}{2} f_Z(y+1) = \begin{cases} \frac{1}{4}, & -2 \leq y < 0, \\ \frac{1}{4} + \frac{1}{4}, & y = 0 \\ \frac{1}{4}, & 0 < y \leq 2 \end{cases}
\end{aligned}$$

(The different value of the density at $y = 0$ doesn't affect the conditional entropy, or any other meaningful property, since the distribution is continuous.) Therefore $\mathcal{C} = \max_{P_V} I(V;Y) = 1$ bit, and $P(V = -1) = P(V = 1) = 0.5$ is a capacity-achieving input distribution.

(c) This part is from Examples Paper 3. Let the density of X be f. Consider the relative entropy between f and ϕ, where ϕ is the $\mathcal{N}(0, \tau^2)$ density, i.e., $\phi(x) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-x^2/2\tau^2}$. We have [20%]

$$\begin{aligned}
0 \leq D(f\|\phi) &= \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{\phi(x)} \, dx \\
&= \int_{-\infty}^{\infty} f(x) \log f(x) - \int_{-\infty}^{\infty} f(x) \left[-\log \sqrt{2\pi\tau^2} - \frac{x^2}{2\tau^2} \log e \right] \, dx \\
&= \log \sqrt{2\pi\tau^2} \int_{-\infty}^{\infty} f(x) \, dx \; + \; \frac{\log e}{2\tau^2} \int_{-\infty}^{\infty} x^2 f(x) \, dx \; + \; \int_{-\infty}^{\infty} f(x) \log f(x) \\
&\overset{(a)}{\leq} \frac{1}{2} \log 2\pi\tau^2 \; + \; \frac{\log e}{2\tau^2} \tau^2 - h(X) = \frac{1}{2} \log 2\pi e \tau^2 - h(X).
\end{aligned}$$

(a) holds because $\mathbb{E}X^2 = \int_{-\infty}^{\infty} x^2 f(x) \, dx \leq \tau^2$. This proves that $h(X) \leq \frac{1}{2} \log 2\pi e \tau^2$. Both inequalities in the chain above become equalities when $D(f\|\phi) = 0$, i.e., when $f = \phi$.

(d) i) Since $\mathrm{Var}(X_1) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = E\Big((X_1 - \mathbb{E}[X_1])^2\Big)$ must be non-negative, we must have $\mathbb{E}[X_1^2] \geq (\mathbb{E}[X_1])^2$, or $P \geq \mu^2$. [10%]

ii) We have

$$\mathbb{E}[Y^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\mathbb{E}[X_1X_2] + 4\mu^2 - 4\mu(\mathbb{E}X_1 + \mathbb{E}X_2) = 2P - 2\mu^2,$$

where last equality is obtained by using $\mathbb{E}[X_1^2] = \mathbb{E}[X_2^2] = P$ and the independence of X_1 and X_2 which implies $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = \mu^2$. Using part (c), we then have $h(Y) \leq \frac{1}{2}\log_2(2\pi e(2P - 2\mu^2))$, with equality when $X_1 \sim \mathcal{N}(\mu, P)$ and $X_2 \sim \mathcal{N}(\mu, P)$ independent of each other.

ii) When X_1, X_2 can be dependent,

$$\mathbb{E}[Y^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\mathbb{E}[X_1X_2] - 4\mu^2 = 2P + 2\mathbb{E}[X_1X_2] - 4\mu^2.$$

The correlation $\mathbb{E}[X_1X_2]$ is maximized when $X_1 = X_2$, in which case $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1^2] = P$. Therefore $\mathbb{E}[Y^2] \leq 4(P - \mu^2)$, with equality when $X_1 = X_2$. Using part (b) again, we have $h(Y) \leq \frac{1}{2}\log_2(2\pi e(4P - 4\mu^2))$, with equality when $X_1 \sim \mathcal{N}(\mu, P - \mu^2)$ and $X_2 = X_1$. [15%]

Question 4

(a) Since the $\mathbf{G}$ is an $n \times k$ matrix, we have $n = 6$ and $k = 3$. The dimension $k = 3$ and the rate is $R = k/n = 1/2$. [10%]

(b) Denoting the information bits by $\underline{x} = [x_1, x_2, x_3]$ the codeword $\underline{c} = \underline{x} \cdot \mathbf{G} = [1, c_2, c_3, c_4, 0, 1]$. Using the given $\mathbf{G}$, we obtain [15%]

$$[x_1,\ x_2,\ x_2 + x_3,\ x_1 + x_2 + x_3,\ x_1 + x_3,\ x_1 + x_2] = [1, c_2, c_3, c_4, 0, 1].$$

Solving for x_1, x_2, x_3 from the three equations with non-erased bits, we get $x_1 = 1$, $x_2 = 0$, $x_3 = 1$. This gives $c_2 = 0, c_3 = 1, c_4 = 0$.

(c) Replacing the second row of $\mathbf{G}$ by the sum of the second and third rows, we get a systematic generator matrix: [10%]

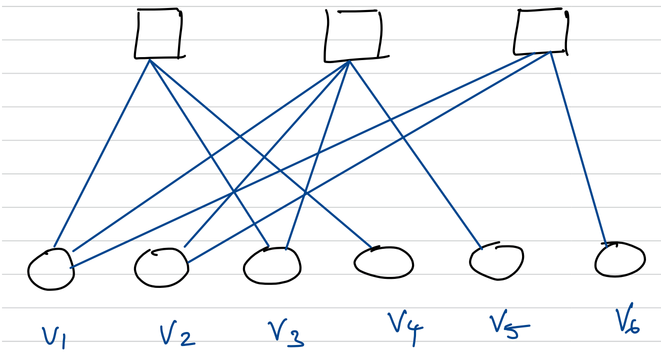
$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} := G_{\text{sys}}$$

(d) For systematic $G_{\text{sys}} = [\mathbf{I}_k \mid \mathbf{P}]$, the parity check matrix $\mathbf{H} = [\mathbf{P}^T \mid \mathbf{I}_{n-k}]$. Therefore: [10%]

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

(e) The minimum distance of the code is the smallest number of columns in $\mathbf{H}$ that add up to $\mathbf{0}$. Columns 2, 5, and 6 add to $\mathbf{0}$, and there are no two columns that add to to $\mathbf{0}$. Therefore the minimum distance is 3. [10%]

(f) The factor graph corresponding to the parity check matrix is: [10%]



(g) The received sequence is $\underline{y} = [-0.6,\ 0.5,\ 0.2,\ 0.4\ -1.2,\ 0.7]$.

i) The message transmitted by variable node j in the first iteration is $L(y_j) = 2y_j/\sigma^2$, for $j = 1, \ldots, 6$. Therefore, the message sent by the first v-node is $L(y_1) = -1.2$. [5%]

ii) The third code bit is connected to the first and second check nodes. Therefore its LLR after one complete round of message passing is:

$$L_3 = L(y_3) + L_{c_1 \to v_3} + L_{c_2 \to v_3},$$

where $L_{c_1\to v_3}, L_{c_2\to v_3}$ are the messages sent from check nodes 1 and 2, respectively, to v_3. The message $L_{c_1\to v_3}$ is determined by the initial LLRs received from v_1 and v_4 (the v-nodes other than v_3 connected to the third check node):

$$L_{c_1\to v_3} = 2\tanh^{-1}\left[\tanh(L(y_1)/2)\tanh(L(y_4)/2)\right] = 2\tanh^{-1}\left[\tanh(-0.6)\tanh(0.4)\right] = -0.4139.$$

[30%]

Similarly,

$$\begin{aligned} L_{c_2\to v_3} &= 2\tanh^{-1}\left[\tanh(L(y_1)/2)\tanh(L(y_2)/2)\tanh(L(y_5)/2)\right] \\ &= 2\tanh^{-1}\left[\tanh(-0.6)\tanh(0.5)\tanh(-1.2)\right] = 0.4199. \end{aligned}$$

Using the above along with $L(y_3) = 0.2$, we obtain:

$$L_3 = L(y_3) + L_{c_1\to v_3} + L_{c_2\to v_3} = 0.4 - 0.4139 + 0.4199 = 0.406$$

The sixth code bit is connected only to the third check node. Therefore its LLR after one complete round of message passing is: $L_6 = L(y_6) + L_{c_3\to v_6}$. The message $L_{c_3\to v_6}$ is determined by the initial LLRs received from v_1 and v_2.

$$L_{c_3\to v_6} = 2\tanh^{-1}\left[\tanh(L(y_1)/2)\tanh(L(y_2)/2))\right] = 2\tanh^{-1}\left[\tanh(-0.6)\tanh(0.5)\right] = -0.5069.$$

Using the above along with $L(y_6) = 1.4$, we obtain:

$$L_6 = L(y_6) + L_{c_3\to v_6} = 1.4 - 0.5069 = 0.8931.$$

Since $L_3 > 0$ and $L_6 > 0$, both code bits will be decoded to 0.