Problem 1 – Information Measures and Source Coding (30 Points)

(a) **(10 Points)** Consider a source X on the alphabet $\mathcal{X} = \{1, 2, 3\}$ having probabilities

$$P_X(1) = \frac{1}{2}, \quad P_X(2) = p, \qquad P_X(3) = \frac{1}{2} - p$$

for some $p \in \left(0, \frac{1}{2}\right)$. For which value(s) of $p \in \left(0, \frac{1}{2}\right)$ will the Shannon-Fano code give the same average codeword length as the Huffman code? Explain.

Solution. *The Huffman code assigns length* 1 *to symbol* 1 *and length* 2 *to the other symbols (e.g.,* 0*,* 10*,* 11*). The Shannon-Fano code also lets symbol* 1 *have length* 1*, but it only assigns both others to have length* 2 *if* $p = \frac{1}{4}$*. This is because:*

- *If* $p \in \left(0, \frac{1}{4}\right)$ *then* $\left\lceil \log_2 \frac{1}{p} \right\rceil \geq 3$*, but also* $\frac{1}{2} - p < \frac{1}{2}$ *and hence* $\left\lceil \log_2 \frac{1}{\frac{1}{2}-p} \right\rceil \geq 2$*. So the average code length is higher due to the length-*3 *(or more) codeword.*
- *Similarly, if* $p \in \left(\frac{1}{4}, \frac{1}{2}\right)$ *then* $\left\lceil \log_2 \frac{1}{\frac{1}{2}-p} \right\rceil \geq 3$ *and* $\left\lceil \log_2 \frac{1}{p} \right\rceil \geq 2$ *(by the same argument as above, with the roles of* p *and* $\frac{1}{2} - p$ *switched).*

Hence, the answer is $p = \frac{1}{4}$ *only, as in this case both* $x = 2$ *and* $x = 3$ *have probability* $\frac{1}{4}$ *and hence Shannon-Fano length* $\log_2 4 = 2$*.*

(b) **(10 Points)** Consider a length-n source sequence $\mathbf{X} \in \mathcal{X}^n$ whose probability mass function (PMF) is given by

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2}\left(\prod_{i=1}^{n} P_1(x_i)\right) + \frac{1}{2}\left(\prod_{i=1}^{n} P_2(x_i)\right),$$

for some PMFs P_1 and P_2 on the alphabet $\mathcal{X}$, where $\mathbf{x} = (x_1, \ldots, x_n)$. Describe a fixed-length source encoding and decoding method for $\mathbf{X}$ based on typical sets, and briefly explain why it can reconstruct $\mathbf{X}$ with probability arbitrarily close to one for any rate $R > \max\{H(P_1), H(P_2)\}$, where $H(P)$ denotes the entropy of a random variable with PMF given by P*. (Note: You may use any properties of typical sets that we proved in class.)*

Solution. *Let* $\mathcal{T}_n^{(1)}(\epsilon)$ *and* $\mathcal{T}_n^{(2)}(\epsilon)$ *be the typical sets for* P_1 *and* P_2*. Consider a source coding scheme that assigns every sequence from* $\mathcal{T}_n^{(1)}(\epsilon) \cup \mathcal{T}_n^{(2)}(\epsilon)$ *a unique index in* $\{1, \ldots, M-1\}$*, and assigns all other sequences to* M*.*

Hence, if $\mathbf{X} \in \mathcal{T}_n^{(1)}(\epsilon) \cup \mathcal{T}_n^{(2)}(\epsilon)$*, then the decoder can recover* $\mathbf{X}$*. Observe that* $P_{\mathbf{X}}$ *draws from* $\prod_{i=1}^n P_1(x_i)$ *half the time and from* $\prod_{i=1}^n P_2(x_i)$ *half the time. In the former (respectively, latter) case, the vector lies in* $\mathcal{T}_n^{(1)}(\epsilon)$ *(respectively,* $\mathcal{T}_n^{(2)}(\epsilon)$*) with probability arbitrarily close to one. Hence, overall* $\mathbf{X} \in \mathcal{T}_n^{(1)}(\epsilon) \cup \mathcal{T}_n^{(2)}(\epsilon)$ *with probability arbitrarily close to one.*

It only remains to characterize the number of sequences needed:

$$\begin{aligned} M &\leq |\mathcal{T}_n^{(1)}(\epsilon)| + |\mathcal{T}_n^{(2)}(\epsilon)| + 1 \\ &\leq 2^{n(H(P_1)+\epsilon)} + 2^{n(H(P_2)+\epsilon)} + 1 \\ &\leq 3 \times 2^{n(\max\{H(P_1),H(P_2)\}+\epsilon)}. \end{aligned}$$

Since $M = 2^{nR}$ *and* ϵ *may be arbitrarily small, it follows that any* $R > \max\{H(P_1), H(P_2)\}$ *suffices (the factor of* 3 *is negligible for large* n*).*

(c) **(10 Points)** Consider the discrete random variable $X \sim \text{Bernoulli}\left(\frac{1}{2}\right)$ (i.e., $P_X(0) = P_X(1) = \frac{1}{2}$) and the continuous random variable $Z \sim \text{Uniform}[0,1]$ (independent of X), and let $Y = \alpha X + Z$ for some $\alpha > 0$. Show that $h(Y) = \min\{1, \alpha\}$, where $h(\cdot)$ is the differential entropy.

Solution. *If $\alpha \geq 1$, then $Y = \alpha X + Z$ is uniform over $[0,1] \cup [\alpha, \alpha+1]$. This set has a total length of 2, so the probability density function must equal $f_Y(y) = \frac{1}{2}$ within this set. This gives $h(Y) = \mathbb{E}\big[\log_2 \frac{1}{f_Y(Y)}\big] = \log_2 2 = 1$.*

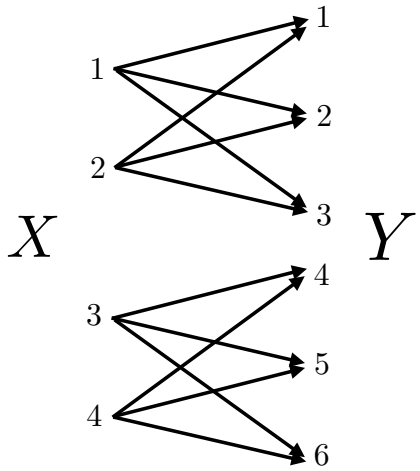
If $\alpha \in (0,1)$, then Y is uniform on $[0,1]$ given $X = 0$, and uniform on $[\alpha, 1+\alpha]$ given $X = 1$. Hence, and since $P_X(0) = P_X(1) = \frac{1}{2}$, the density equals $f_Y(y) = \frac{1}{2}$ for $y \in (0,\alpha) \cup (1, 1+\alpha)$ (i.e., y values that only one value $x \in \{0,1\}$ can lead to), and $f_Y(y) = 1$ for $y \in (\alpha, 1)$ (i.e., y values that both values $x \in \{0,1\}$ can lead to).

When evaluating $\mathbb{E}\big[\log_2 \frac{1}{f_Y(Y)}\big]$, the values with $f_Y(Y) = 1$ give $\log_2 \frac{1}{f_Y(Y)} = 0$, whereas the values with $f_Y(Y) = \frac{1}{2}$ give $\log_2 \frac{1}{f_Y(Y)} = \log_2 2 = 1$. The probability of the latter case is $\frac{1}{2}$ times the total length of $(0,\alpha) \cup (1, 1+\alpha)$, which is $\frac{1}{2} \times 2\alpha = \alpha$.

Combining these two cases gives $h(Y) = \min\{1, \alpha\}$.

Problem 2 – Channel Coding I (30 Points)

(a) **(8 Points)** Consider a discrete memoryless channel of the following form with input alphabet $\mathcal{X} = \{1,2,3,4\}$ and output alphabet $\mathcal{Y} = \{1,2,3,4,5,6\}$:



The precise transition probabilities are not shown, but any absent transitions correspond to a conditional probability of zero.

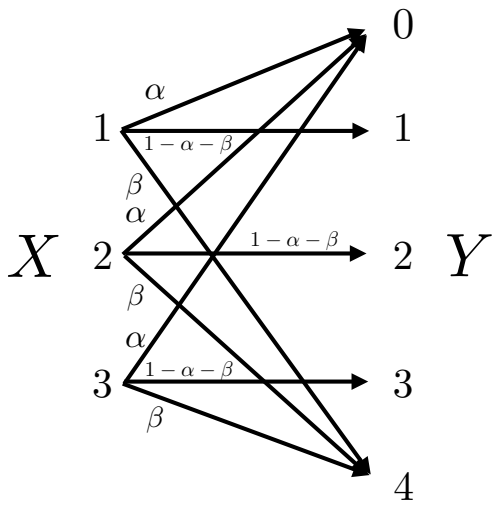
(i) In at most two sentences, explain why the channel capacity is at least 1 bit/use.

(ii) In at most two sentences, explain why the channel capacity cannot exceed 2 bits/use.

Solution. *(i) 1 bit/use can be attained with zero error probability by sending uncoded information using inputs 1 and 3.*

(ii) Since the input alphabet size is 4, we have $C = \max_{P_X} I(X;Y) \leq \max_{P_X} H(X) = 2$ bits/use (uniform maximizes entropy).

(b) **(12 Points)** Consider the following 3-input 5-output discrete memoryless channel, with parameters $\alpha \in \left(0, \frac{1}{2}\right)$ and $\beta \in \left(0, \frac{1}{2}\right)$:



Using the capacity formula $C = \max_{P_X} I(X;Y)$, prove that the capacity is

$$C = (1 - \alpha - \beta) \log_2 3 \quad \text{bits/use.}$$

Solution. *Fix P_X and write*

$$I(X;Y) = H(X) - H(X|Y) = H(X) - \sum_y P_Y(y) H(X|Y = y).$$

Observe that for $y \in \{1, 2, 3\}$ we have $H(X|Y = y) = 0$, since there is only one x that could have led to any such y. On the other hand, for $y = 0$, we have

$$P_{X|Y}(x|0) = \frac{P_{XY}(0, y)}{P_Y(0)} = \frac{P_X(x)\alpha}{\alpha} = P_X(x),$$

where we used $P_Y(0) = \sum_x P_X(x) P_{Y|X}(0|x) = \sum_x P_X(x)\alpha = \alpha$. By the same argument, we have

$$P_{X|Y}(x|4) = P_X(x)$$

and $P_Y(4) = \beta$. Combining these gives

$$\sum_y P_Y(y) H(X|Y = y) = \alpha H(X) + \beta H(X),$$

and substitution into the first equation above gives

$$I(X;Y) = (1 - \alpha - \beta) H(X).$$

Then, since uniform maximizes entropy, we have $H(X) \leq \log_2 3$ with equality if $P_X = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, so the capacity is $C = (1 - \alpha - \beta) \log_2 3$.

(c) **(10 Points)** Consider a continuous memoryless channel described by $Y = X + Z$, where X is subject to a power constraint $\mathbb{E}[X^2] \le P$, but in contrast with the AWGN channel in the lecture, the noise is correlated with the input: $Z = \alpha X + V$ for some $\alpha > 0$, where $V \sim N(0, \sigma^2)$ is independent of X. Starting with the formula

$$C = \max_{f_X : \mathbb{E}[X^2] \le P} I(X;Y),$$

find the channel capacity in terms of P, σ^2, and α.

(Note: Recall that the differential entropy of an $N(0,\sigma^2)$ random variable is $\frac{1}{2}\log_2(2\pi e\sigma^2)$)

Solution. *Fix f_X satisfying $\mathbb{E}[X^2] \le P$, and observe that $(X,Y) \sim f_X \times f_{Y|X}$ satisfies the following:*

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h((1+\alpha)X + V|X) \\ &= h(Y) - h(V|X) \\ &= h(Y) - h(V), \end{aligned}$$

where the four lines use (i) definition of mutual information, (ii) definition of Y, (iii) shifting by a constant doesn't change the entropy (and $(1+\alpha)X$ is constant given X), (iv) V is independent of X.

Since $V \sim N(0,\sigma^2)$, we have $h(V) = \frac{1}{2}\log_2(2\pi e\sigma^2)$. In addition, since $Y = X + Z = (1+\alpha)X + V$, and X and V are independent, we have $\mathsf{Var}[Y] = (1+\alpha)^2 P + \sigma^2$, so the property of Gaussian maximizing entropy gives

$$h(Y) \le \frac{1}{2}\log_2\left(2\pi e((1+\alpha)^2 P + \sigma^2)\right).$$

Substitution into $I(X;Y) = h(Y) - h(V)$ gives

$$I(X;Y) \le \frac{1}{2}\log_2 \frac{(1+\alpha)^2 + \sigma^2}{\sigma^2} = \frac{1}{2}\log_2\left(1 + \frac{(1+\alpha)^2 P}{\sigma^2}\right).$$

Equality holds (following from the same in the $h(Y)$ upper bound) if $X \sim N(0,P)$ (which implies $Y \sim N(0,(1+\alpha)^2 P + \sigma^2)$), so we conclude that $C = \frac{1}{2}\log_2\left(1 + \frac{(1+\alpha)^2 P}{\sigma^2}\right)$.

Problem 3 – Linear Codes (20 Points)

(a) **(8 Points)** Consider the repetition code that maps $0 \to 00000$ and $1 \to 11111$. Interpreting this as a linear code, write down the generator matrix $\mathbf{G}$, the parity check matrix $\mathbf{H}$, and the minimum distance $d_{\min}$. Explanations are not required.

Solution. *We have $\mathbf{G} = [1\ 1\ 1\ 1\ 1]$ and*

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

The minimum distance is $d_{\min} = 5$.

(b) **(12 Points)** Let $\mathbf{G}_1$ (respectively, $\mathbf{G}_2$) be a generator matrix whose code has rate R_1 (respectively, R_2) and minimum distance $d_{\min,1}$ (respectively, $d_{\min,2}$). Suppose that $\mathbf{G}_1$ and $\mathbf{G}_2$ have the same number of rows, and consider the generator matrix

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix}.$$

Consider the linear code with generator matrix $\mathbf{G}$, and answer the following:

(i) Show that the rate R of the code satisfies $R = \frac{1}{(1/R_1)+(1/R_2)}$.

(ii) Show that the minimum distance $d_{\min}$ of the code satisfies $d_{\min} \geq d_{\min,1} + d_{\min,2}$.

(iii) Do there exist any $\mathbf{G}_1$ and $\mathbf{G}_2$ such that $d_{\min} > d_{\min,1} + d_{\min,2}$? Explain.

Solution. *(i) Let k be the number of rows in both $\mathbf{G}_1$ and $\mathbf{G}_2$, and let n_1 and n_2 be the number of columns. Then we have $R_1 = \frac{k}{n_1}$, $R_2 = \frac{k}{n_2}$, and $R = \frac{k}{n_1+n_2}$. Re-arranging the former two equations gives $n_1 = \frac{k}{R_1}$ and $n_2 = \frac{k}{R_2}$, so substitution into the rate equation gives*

$$R = \frac{1}{(1/R_1) + (1/R_2)}.$$

(ii) Let $\mathbf{x} = \mathbf{uG} = [\mathbf{uG}_1 \quad \mathbf{uG}_2]$ and $\mathbf{x}' = \mathbf{u}'\mathbf{G} = [\mathbf{u}'\mathbf{G}_1 \quad \mathbf{u}'\mathbf{G}_2]$ be the two codewords (with corresponding information bits $\mathbf{u}$ and $\mathbf{u}'$) at a minimum distance from each other, and for brevity write these as $\mathbf{x} = [\mathbf{x}_1 \quad \mathbf{x}_2]$ and $\mathbf{x}' = [\mathbf{x}'_1 \quad \mathbf{x}'_2]$. Then we have $d_{\mathrm{H}}(\mathbf{x}, \mathbf{x}') = d_{\min}$ by construction, but we also have

$$d_{\mathrm{H}}(\mathbf{x}, \mathbf{x}') = d_{\mathrm{H}}(\mathbf{x}_1, \mathbf{x}'_1) + d_{\mathrm{H}}(\mathbf{x}_2, \mathbf{x}'_2) \geq d_{\min,1} + d_{\min,2},$$

since $\mathbf{x}_1, \mathbf{x}'_1$ are distinct codewords generated by $\mathbf{G}_1$, and similarly for $\mathbf{x}_2, \mathbf{x}'_2$.

(iii) Yes. Consider the Hamming code:

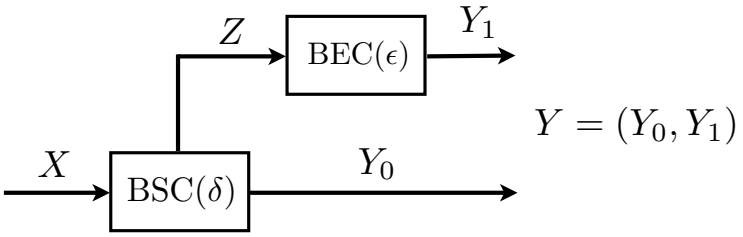
$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Let $\mathbf{G}_1$ be the 4×4 identity matrix, and let $\mathbf{G}_2$ be the 4×3 part on the right. Then $d_{\min,1} = 1$ (e.g., just consider $\mathbf{u} = (0,0,0,0)$ and $\mathbf{u}' = (1,0,0,0)$) and $d_{\min,2} = 0$ (e.g., consider $\mathbf{u} = (0,0,0,0)$ and $\mathbf{u}' = (1,1,1,0)$), but we know that the Hamming code has minimum distance 3.

Problem 4 – Channel Coding II (20 Points)

(a) **(10 Points)** Recall from the lecture that the binary symmetric channel (BSC) has $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and flips the input with probability δ, and that the binary erasure channel (BEC) has $\mathcal{X} = \{0,1\}$ and $\mathcal{Y} = \{0,1,e\}$, and given any input, the output is e with probability ϵ and equals the input otherwise.

Consider the following channel with input X and output $Y = (Y_0, Y_1)$:



Here BEC(ϵ) is a binary erasure channel with parameter $\epsilon \in (0,1)$, and BSC(δ) is a binary symmetric channel with parameter $\delta \in (0,1)$ that not only provides the usual output (written as Y_0 here), but also outputs the value $Z = \mathbf{1}\{X \neq Y_0\}$ equaling 1 if a flip occurred, and 0 otherwise. We can view this channel $P_{Y|X}$ as being a BSC in which we get to "peek" at the noise a fraction $1-\epsilon$ of the time.

Assuming that the "flip event" in the BSC and the "erasure event" in the BEC occur independently of each other, show that the overall channel has capacity $C = 1 - \epsilon H_2(\delta)$, where $H_2(\delta) = \delta \log_2 \frac{1}{\delta} + (1-\delta)\log_2 \frac{1}{1-\delta}$ is the binary entropy function.

Solution. *For the BSC, the flip probability is δ regardless of the value of X. Hence, X and Z are independent. Since Y_1 is produced from Z with no direct dependence on X, it follows that X and Y_1 are also independent. This will be used below.*

Consider the capacity $C = \max_{P_X} I(X;Y) = \max_{P_X} I(X;Y_0,Y_1)$, and note that

$$I(X;Y_0,Y_1) = I(X;Y_1) + I(X;Y_0|Y_1) = I(X;Y_0|Y_1)$$

by the chain rule followed by the independence of X and Y_1. Then, we have

$$\begin{aligned} I(X;Y_0|Y_1) &\overset{(i)}{=} H(Y_0|Y_1) - H(Y_0|X,Y_1) \\ &\overset{(ii)}{\leq} 1 - H(Y_0|X,Y_1) \\ &\overset{(iii)}{=} 1 - (1-\epsilon)0 + \epsilon H_2(\delta) \\ &= 1 - \epsilon H_2(\delta), \end{aligned}$$

where (i) uses the definition of mutual information, (ii) uses the fact the the entropy of a binary random variable is at most one, and (iii) uses the fact that there is no uncertainty in Y_0 given (X,Y_1) when $Y_1 \neq e$ (erasure) but there is uncertainty $H_2(\delta)$ when $Y_1 = e$ (since X is known but passed through BSC(δ)*, whereas Y_1 carries no information due to being erased).*

To show that $C = 1 - \epsilon H_2(\delta)$, it only remains to show that there exists P_X such that inequality (ii) holds with equality. We claim that $P_X = \left(\frac{1}{2}, \frac{1}{2}\right)$ suffices, i.e., gives $H(Y_0|Y_1) = 1$. This is because $Y_0 = X \oplus Z$ (mod-2 addition), and Y_1 only carries information about Z (it is independent of X as shown above). But even if Z were fully known, the uniformity of X still makes $X \oplus Z$ uniform, meaning that $H(Y_0|Y_1) = 1$.

(b) **(10 Points)** Consider a channel with input $X = (X_A, X_B, X_C)$ and output $Y = (Y_A, Y_B, Y_C)$, where X_A, X_B, X_C, Y_A, Y_B, and Y_C are all binary-valued, with values $\{0,1\}$. The channel transition probabilities are described as follows:

$$(Y_A, Y_B, Y_C) = \begin{cases} (X_A, X_B, X_C) & \text{with probability } 1-\delta \\ (1-X_A, 1-X_B, 1-X_C) & \text{with probability } \delta. \end{cases}$$

That is, either all 3 bits are flipped or none of them are.

Let $\mathcal{C}_1$ be a codebook for a binary symmetric channel with rate R_1 and error probability ϵ_1. Describe how this codebook can be used to communicate over the above channel at rate $2 + R_1$ with error probability ϵ_1.

Solution. *Represent the message as $m = (m_0, m_1)$, where m_0 takes one of 4^n values, and m_1 takes one of $M_1 = 2^{nR_1}$ values.*

To transmit (m_0, m_1), let m_0 index a unique length-n sequence $\mathbf{u}$ with each $u_i \in \{1, 2, 3, 4\}^n$, and let m_1 index a codeword $\mathbf{x}_1$ from the BSC codebook $\mathcal{C}_1$. Then, for each channel use $i = 1, \ldots, n$, do the following:

- *Use u_i to select one of the following triplets: (i) $u = 1 \implies (0, 0, 0)$, (ii) $u = 2 \implies (0, 0, 1)$, (iii) $u = 3 \implies (0, 1, 0)$, (iv) $u = 4 \implies (0, 1, 1)$.*
- *If $x_{1,i} = 0$, transmit the triplet directly. Otherwise, if $x_{1,i} = 0$, flip all 3 bits of the triplet and then transmit.*

The decoder can obtain each u_i with zero probability of error, by just checking whether all 3 bits are the same (1), only the first two are the same (2), only the first and third are the same (3), or only the last two are the same (4). In addition, $x_{1,i}$ is simply equal to the first bit in the triplet, so taking the bit of each output triplet gives $\mathbf{Y}_1$, corresponding to passing $\mathbf{X}_1$ through a binary symmetric channel. This can be decoded with error probability ϵ_1 by assumption.

Finally, the number of messages is $M = 4^n \cdot 2^{nR_1} = 2^{n(2+R_1)}$, so the rate is $R = 2 + R_1$.

END OF PAPER