1. Rate-Distortion (40 points)

Consider the $M$-ary alphabet $\mathcal{Z} = \{0, 1, \ldots, M-1\}$ and the function $\rho$ satisfying $\rho(z) \geq 0$ for all $z \in \mathcal{Z}$, with equality if and only if $z = 0$.

(a) (10 points) For parameter $\beta \geq 0$ let $Z_\beta$ be a $\mathcal{Z}$-valued random variable with PMF

$$P(Z_\beta = z) = \frac{e^{-\beta \rho(z)}}{\sum_{z' \in \mathcal{Z}} e^{-\beta \rho(z')}},$$

$z \in \mathcal{Z}$.

Show that any $\mathcal{Z}$-valued random variable $Z$ for which

$$E[\rho(Z)] \leq E[\rho(Z_\beta)]$$

must satisfy

$$H(Z) \leq H(Z_\beta),$$

with equality if and only if $Z$ has the same distribution as $Z_\beta$.

Hint: Note that $0 \leq D(P_Z \parallel P_{Z_\beta})$ and write out the right hand side of this inequality.

(b) (5 points) Consider the function

$$\phi(D) = \max_{P_Z: E[\rho(Z)] \leq D} H(Z), \quad D \geq 0,$$

where the maximization is over all $\mathcal{Z}$-valued random variables satisfying the indicated constraint. Find $\phi(D_\beta)$, where $D_\beta = E[\rho(Z_\beta)]$, i.e., find

$$\phi(D_\beta) = \max_{P_Z: E[\rho(Z)] \leq D_\beta} H(Z).$$

(c) (5 points) Show that

$$\left| \left\{ z^n \in \mathcal{Z}^n : \frac{1}{n} \sum_{i=1}^{n} \rho(z_i) \leq D \right\} \right| \leq 2^{n\phi(D)}.$$

(d) (10 points) Let $R(D)$ be the rate distortion function for an $M$-ary source $U$, with the same $M$-ary reconstruction alphabet and distortion measure given by

$$d(u, v) = \rho(u - v),$$

where subtraction here is modulo-$M$. That is:

$$R(D) = \min_{p(v|u): E[\rho(U-V)] \leq D} I(U; V), \quad D \geq 0.$$

Show that

$$R(D) \geq H(U) - \phi(D). \quad (1)$$
(e) *(5 points)* Suppose $U$ is uniformly distributed over \(\{0, 1, \ldots, M - 1\}\). Show that in this case, eq. (1) holds with equality, i.e.,

\[
R(D) = \log M - \phi(D), \quad D \geq 0.
\]

(f) *(5 points)* Give a qualitative “geometric” explanation to inequality (1) via the result of Part (c).

**Solution:**

(a) Analogously as we did to establishing that the distribution that maximizes the entropy for the continues case is the Gaussian distribution, we have:

\[
0 \leq D(P_Z||P_{Z_\beta}) = \sum_z P_Z(z) \log \frac{P_Z(z)}{P_{Z_\beta}(z)} = -H(Z) + E_Z \left[ \frac{\beta \rho(Z)}{\ln(2)} + \log \sum_{z'} e^{-\beta \rho(z')} \right] \leq -H(Z) + E_Z \left[ \frac{\beta \rho(Z_\beta)}{\ln(2)} + \log \sum_{z'} e^{-\beta \rho(z')} \right] = -H(Z) + E \left[ \frac{1}{P_{Z_\beta}} \log \right] = -H(Z) + H(Z_\beta)
\]

and we see that the two inequalities hold with equality if and only if $P_Z = P_{Z_\beta}$.

(b) It follows immediately from the previous part that $\max_{1 \leq \rho(Z) \leq D_\beta} H(Z)$ is uniquely attained by $Z_\beta$.

Hence:

\[
\phi(D_\beta) = H(Z_\beta) = E \left[ \frac{\beta \rho(Z)}{\ln(2)} + \log \sum_{z'} e^{-\beta \rho(z')} \right] = \frac{\beta D_\beta}{\ln(2)} + \log \sum_{z'} e^{-\beta \rho(z')}
\]

(c)

\[
\left| \left\{ z^n \in \mathcal{Z}^n : \frac{1}{n} \sum_{i=1}^{n} \rho(z_i) \leq D \right\} \right| = \sum_{P \in P_n, \sum_{z} P(z) \rho(z) \leq D} |T(p)|
\]
\[(n + 1)^{nH'} \leq \sum_{P \in \mathbb{P}, \sum_{z} P(z) \rho(z) \leq D} |T(p)| \leq |\mathbb{P}| 2^{nH'},\]

where

\[H' = \max_{P \in \mathbb{P}, \sum_{z} P(z) \rho(z) \leq D} H(P)\]

And notice that as \(n\) goes to infinity, \(H'\) goes to \(\phi(D)\).

The result follows since both left and right sides of the inequality are \(\cdot 2^{n\phi(D)}\).

(d) Similar to what we have done in the calculation of \(R(D)\) for both binary and Gaussian source:

If \(U, V\) are in the feasible set for the min in (1d), then

\[I(U, V) = H(U) - H(U|V)\]

\[\geq H(U) - H(U - V)\]

\[\geq H(U) - \phi(D)\]

where the last inequality follows because \(\mathbb{E}[\rho(U - V)] \leq D\).

(e) Again, analogously as in the binary and Gaussian case, we have that if \(U\) is uniform, we can find a joint distribution in the feasible set that achieves equality in both the inequalities of previous part.

Specifically, we have \(U = V + Z(D)\), where \(Z(D)\) is the achiever of the max defining \(\phi(D)\). Both \(U\) and \(V\) are uniform.

(f) Again, analogous to the cases we have seen, let \(V^n\) be a reconstruction in the codebook. Consider the "distortion ball of radius \(D\) around \(V^n\), or in other words, the set of source sequences covered by \(V^n\) to within distortion \(D\):

Define \(B(D, V^n) = \{u^n : \rho(u^n - V^n) \leq D\}\), where \(\rho(Z^n) = \frac{1}{n} \sum_{i=1}^{n} \rho(Z_i)\).

Now, by part (c), we have:

\[|B(D, V^n)| = |\{z^n | \rho(z^n) \leq D\}| \doteq 2^{n\phi(D)}\]

If we want to achieve distortion no more than \(D\), we need to cover the typical set with distortion \(D\) balls around the reconstructions. How many such balls will we need? At least \((\text{size of typical set}) / (\text{size of ball})\), i.e. \(2^{nH(U)} / 2^{n\phi(D)} = 2^{n(H(U) - \phi(D))}\).

2. List Codes (30 points)

A \((2^{nR}, 2^{nL}, n)\) list code for a DMC \(p(y|x)\) with capacity \(C\) consists of an encoder that assigns a codeword \(x^n(m)\) to each message \(m \in \{1, \ldots, 2^{nR}\}\) and a decoder that, upon receiving \(y^n\), outputs a list of messages \(L(y^n) \subset \{1 \ldots 2^{nR}\}\) of size \(|L| \leq 2^{nL}\). The decoder’s goal is to output a list that contains the transmitted message. An error occurs if the list does not contain the transmitted message \(M\), i.e., the probability of error is \(P_e(n) = P\{M \notin L(Y^n)\}\) (where the message \(M\) is uniformly distributed on \(\{1, \ldots, 2^{nR}\}\)). A rate-list pair \((R, L)\) is said to be achievable if there exists a sequence of \((2^{nR}, 2^{nL}, n)\) list codes with \(P_e(n) \to 0\) as \(n \to \infty\).
(a) \((10\text{ points})\) Show that if \(R < C + L\) then \((R, L)\) is achievable.

Hint: Verify the following random coding scheme has vanishing probability of error.

- Random codebook generation as in the standard achievability.
- Given \(y^n, L(y^n)\) will be the list of messages \(m\) such that \((X^n(m), y^n)\) are jointly typical, provided the size of this list does not exceed \(2^{nL}\). Otherwise, can choose \(L(y^n)\) arbitrarily to be any list of size not exceeding \(2^{nL}\).
- When establishing that average \(P_e^{(n)}\) (across the ensemble of random codebooks) is vanishing, you can use the fact, due to markov’s inequality, that for any random set \(A\)

\[
P(|A| > 2^{nL}|M = 1) \leq \frac{E[|A||M = 1]}{2^{nL}}
\]

(b) \((10\text{ points})\) Show that for any scheme: 
\[
H(M|Y^n) \leq 1 + nL + P_e^{(n)}nR.
\]

(c) \((10\text{ points})\) Show that for every sequence of \((2^{nR}, 2^{nL}, n)\) list codes with \(P_e^{(n)} \to 0\) as \(n \to \infty\), we have \(R \leq C + L\).

Solution: List Codes

(a) Consider the random codebook generation as in the standard achievability proof for the DMC. For the decoding, let \(A = \{m : (x^n(m); y^n) \in T^{(n)}\}\) and declare \(L = A\) if \(|A| \leq 2^{nL}\) (and take an arbitrary \(L\) otherwise). Suppose \(M = 1\). Then the probability of error is bounded as

\[
P_e^{(n)} = Pr(M \notin L(Y^n)|M = 1) \\
\leq Pr((X^n(1), Y^n) \notin T^{(n)}_e|M = 1) + Pr(|A| > 2^{nL}|M = 1)
\]

By the LLN, the first term tends to zero as \(n \to \infty\). By the Markov inequality and the joint typicality lemma, the second term is bounded as

\[
P_r(|A| > 2^{nL}|M = 1) \leq \frac{E[|A||M = 1]}{2^{nL}} \\
\leq \frac{1}{2^{nL}} \left(1 + \sum_{m=2}^{2^n R} Pr((X^n(m), Y^n) \in T^{(n)}_e)\right) \\
\leq 2^{-nL} + 2^{(R-L-I(X;Y)+\delta)}
\]

which tends to zero as \(n \to \infty\) if \(R < L + I(X;Y)\).

Alternatively, we can partition \(2^{nR}\) messages together into \(2^{n(R-L)}\) equal-size groups and map each group into a single codeword. The encoder sends the group index \(k \in [1 : 2n(R-L)]\) by transmitting \(x^n(k)\) for \(m \in [(k-1)2^{nL} + 1 : k2^{nL}]\). The decoder finds the correct group index \(\hat{k}\) and simply forms the list of messages associated with \(\hat{k}\), i.e., \(L = [(\hat{k}-1)2^{nL} + 1 : \hat{k}2^{nL}]\). Finally, by the channel coding theorem for the standard DMC, the group index can be reliably decoded if \(R - L < C\), which completes the proof of achievability.
(b) Many of you just applied the original Fano’s inequality, which can not be directly applied to this setting. The reason is that we defined the error event in different way. The intuition behind this problem is that given the output $Y^n$, we have a list which might contain the true message with high probability, therefore, the remaining uncertainty will be $\log 2^n L$. Here is more rigorous proof. Let $E$ be a binary random variable that takes value 1 if there is an error, and 0 otherwise.

\[
H(M|Y^n) = H(E|M, Y^n) + H(M|E, Y^n) \\
\leq H(E) + H(M|E, Y^n) \\
\leq 1 + P(E = 0)H(M|E = 0, Y^n) + P(E = 1)H(M|E = 1, Y^n) \\
\leq 1 + (1 - P_e(n)) \log 2^n L + P_e(n) \log 2^n R \\
\leq 1 + nL + P_e(n)nR
\]

(c) Let $\epsilon_n = R P_e(n) + \frac{1}{n}$, where $\epsilon_n \to 0$ as $n \to 0$. Then, $H(M|Y^n) \leq nL + n\epsilon_n$. Therefore, given any sequence of $(2^n R, 2^n L, n)$ codes with $P_e(n) \to 0$ as $n \to \infty$, we have

\[
nR = H(M) \\
= I(M; Y^n) + H(M|Y^n) \\
\leq I(M; Y^n) + nL + n\epsilon_n \\
= \sum_{i=1}^{n} I(X_i; Y_i) + nL + n\epsilon_n \\
= nC + nL + n\epsilon_n
\]

Thus, $R \leq C + L$.

3. Zero-Rate Communication (30 points)

(a) (10 points) Suppose we want to communicate the one-bit message $M \in \{0, 1\}$ through the BSC($p$), $p \leq \frac{1}{2}$, with $n$ channel uses. Assuming $P(M = 0) = P(M = 1) = 1/2$, the most natural idea is to send, say, all zeros if $M = 0$ and all ones if $M = 1$. The decoder will decide $\hat{M} \in \{0, 1\}$ based on the output $Y^n$ using majority vote: if $Y^n$ has more zeros than ones then $\hat{M} = 0$, otherwise $\hat{M} = 1$. Find the exponential decay rate of the probability of error with the number of channel uses

\[
E_2 = \lim_{n \to \infty} -\frac{1}{n} \log P(M \neq \hat{M}).
\]

Equivalently, find $E_2$ such that

\[
P(M \neq \hat{M}) \leq 2^{-nE_2}.
\]
(b) (10 points) Suppose we now want to transmit the ternary message $M \in \{0, 1, 2\}$ through a BSC($p$), $p \leq \frac{1}{2}$, with $n$ channel uses. The message is uniformly distributed on $\{0, 1, 2\}$. We use the following encoding:

- If $M = 0$, we send all zeros.
- If $M = 1$, we send all ones for the first $\frac{2}{3}$ channel uses, and send zeros in the rest of the time.
- If $M = 2$, we send zeros for the first $\frac{1}{3}$ channel uses, and send ones in the rest of the time.

In other words, the codebook is:

$$X^n(0) = 00 \cdots 00 00 \cdots 00$$

$$X^n(1) = 11 \cdots 11 11 \cdots 00$$

$$X^n(2) = 00 \cdots 00 11 \cdots 11$$

The decoder uses “minimum distance decoding” which finds $\hat{M}$ that minimizes the distance between $X^n(\hat{M})$ and $Y^n$, i.e.,

$$\hat{M} = \arg \min_{m \in \{0, 1, 2\}} \sum_{i=1}^{n} |X_i(m) - Y_i|.$$  

Find

$$E_3 = \lim_{n \to \infty} \frac{1}{n} \log P(M \neq \hat{M}).$$

Equivalently, find $E_3$ such that

$$P(M \neq \hat{M}) = 2^{-nE_3}.$$  

(c) (10 points) Suppose now the message is quaternary $M \in \{0, 1, 2, 3\}$, uniformly distributed on $\{0, 1, 2, 3\}$. Suggest a scheme (encoding and decoding) that achieves the same exponential decay rate of the probability of error as that in the previous part. Prove the scheme indeed achieves this exponential decay rate.

Solution: Zero Rate Communication

(a) Let $Z^n = \{Z_1, \cdots, Z_n\}$ be i.i.d. Bern($p$) random variables and $Y_i = X_i \oplus Z_i$. An error occurs if and only if more than half of $Z^n$ is one. In HW7 problem 8, we have seen that

$$Pr \left( \frac{1}{n} \sum_{i=1}^{n} Z_i \geq \frac{1}{2} \right) = 2^{-nD_r}$$
where
\[ D^* = \min_{P: P(1) \geq \frac{1}{2}} D(P \| \text{Bern}(p)). \]

We can easily show that this minimum is achieved by \( P(0) = P(1) = \frac{1}{2} \) and \( D^* = D(\text{Bern}(\frac{1}{2}) \| \text{Bern}(p)) = \frac{1}{2} \log \frac{1}{2p} + \frac{1}{2} \log \frac{1}{2(1-p)} = \frac{1}{2} \log \frac{1}{4p(1-p)}. \)

(b) Note that
\[
\Pr(\hat{M} \neq 0 | M = 0) \geq \Pr\left( \sum_{i=1}^{n} |X_i(0) - Y_i| \geq \sum_{i=1}^{n} |X_i(1) - Y_i| \left| M = 0 \right. \right)
\]
\[
= \Pr(\text{among the first } 2/3 \text{ fraction of } Z_i \text{'s, at least half of them are ones})
\]
\[
= 2^{-\frac{2}{3}nD^*}
\]

where \( D^* \) remains the same as the previous problem. On the other hand,
\[
\Pr(\hat{M} \neq 0 | M = 0) \leq \Pr\left( \sum_{i=1}^{n} |X_i(0) - Y_i| \geq \sum_{i=1}^{n} |X_i(1) - Y_i| \left| M = 0 \right. \right)
\]
\[
+ \Pr\left( \sum_{i=1}^{n} |X_i(0) - Y_i| \geq \sum_{i=1}^{n} |X_i(2) - Y_i| \left| M = 0 \right. \right)
\]
\[
= 2 \times 2^{-\frac{2}{3}nD^*} = 2^{-\frac{2}{3}nD^*}
\]

Therefore, \( \Pr(\hat{M} \neq 0 | M = 0) \leq 2^{-\frac{2}{3}nD^*} \). The same results hold for \( M = 1, 2 \). Now, we can finally conclude that
\[
P(M \neq \hat{M}) = \frac{1}{3} \Pr(\hat{M} \neq 0 | M = 0) + \frac{1}{3} \Pr(\hat{M} \neq 1 | M = 1) + \frac{1}{3} \Pr(\hat{M} \neq 2 | M = 2)
\]
\[
= 2^{-\frac{2}{3}nD^*}
\]

Thus, \( E_3 = \frac{2}{3} D^* = \frac{1}{3} \log \frac{1}{4p(1-p)}. \)

(c) Using the following encoding scheme and the minimum distance decoding,
\[
X^n(0) = \underbrace{00 \cdots 00}_{n/3} \underbrace{00 \cdots 00}_{n/3} \underbrace{00 \cdots 00}_{n/3}
\]
\[
X^n(1) = \underbrace{11 \cdots 11}_{n/3} \underbrace{11 \cdots 11}_{n/3} \underbrace{11 \cdots 11}_{n/3}
\]
\[
X^n(2) = \underbrace{00 \cdots 00}_{n/3} \underbrace{11 \cdots 11}_{n/3} \underbrace{11 \cdots 11}_{n/3}
\]
\[
X^n(2) = \underbrace{11 \cdots 11}_{n/3} \underbrace{00 \cdots 00}_{n/3} \underbrace{11 \cdots 11}_{n/3}
\]

Note that we can use the same codes as previous problem for \( M = 0, 1, 2 \). Using this encoding strategy and the minimum distance decoding as in (b), one can easily show that we can achieve \( \frac{2}{3} D^* \) for the error exponent, i.e.,
\[
E_4 = \frac{1}{3} \log \frac{1}{4p(1-p)}
\]