Final Exam Solution

1. (True of False) [36]

Solution:

a) False, cosider $g(X) = aX$ for $a \neq 0$, then $h(g(X)) = \log|a| + h(X) > h(X)$.

b) True, because $Y - X - g(X)$ forms a Markov chain.

c) False, let $P_X = Q_X \sim \text{Ber}(0.5)$ and $W^{(1)}_{Y|X}, W^{(2)}_{Y|X}$ be two different Z channels:

$$W^{(1)}_{Y|X}(y|x) = \begin{cases} 1, & x = y = 0, \\ p_1, & x = 1, y = 0, \\ 1 - p_1, & x = 1, y = 1. \end{cases}, W^{(2)}_{Y|X}(y|x) = \begin{cases} 1, & x = y = 0, \\ p_2, & x = 1, y = 0, \\ 1 - p_2, & x = 1, y = 1. \end{cases}, p_1 \neq p_2.$$

Then $P_Y \sim \text{Ber}\left(\frac{1-p_1}{2}\right), Q_Y \sim \text{Ber}\left(\frac{1-p_2}{2}\right), D(P_X||Q_X) = 0 < D(P_Y||Q_Y)$.

d) False, consider $\mathsf{P}_S \sim \text{Ber}(0.9)$, then the most probable sequence $s^* = (1, 1, \cdots, 1)$ with probability $(0.9)^n$. Direct calculation gives

$$-\frac{\log{(0.9)^n}}{n} = -\log 0.9 < -(0.1 \log 0.1 + 0.9 \log 0.9) = \mathrm{H_b}(0.9).$$

Therefore, for small δ, $\mathsf{Pr}\{S^n = s^*\} > 2^{-n(H(S)-\delta)}$ which violates the property of the typical sequence.

e) False, since zero distortion does not imply $\hat{S} = S$.
To see this, let $S \sim \text{Ber}\left(\frac{1}{2}\right)$ and $d(s, \hat{s}) = 0, \forall (s, \hat{s})$. Then $R(0) = 0 < 1 = H(S)$.

Grading Policy:

- Correct true or false: +2
- Reasoning: +6 or +4 (depends on the total score of the subproblem)
- Any typo: −1

2. (Hypothesis Testing with Rejection) [8]

Solution:

$$\pi_r^{(n)}(\phi) = \sum_{x^n \in \mathcal{X}^n} P_0^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = r\} + \sum_{x^n \in \mathcal{X}^n} P_1^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = r\}.$$

$$\pi_e^{(n)}(\phi) = \sum_{x^n \in \mathcal{X}^n} P_0^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = 1\} + \sum_{x^n \in \mathcal{X}^n} P_1^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = 0\}.$$

Hence,

$$\begin{aligned}&\pi_r^{(n)}(\phi) + \pi_e^{(n)}(\phi) \\ &= 2 - \left(\sum_{x^n \in \mathcal{X}^n} P_0^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = 0\} + \sum_{x^n \in \mathcal{X}^n} P_1^{\otimes n}\{X^n = x^n\} \mathbb{1}\{\phi(x^n) = 1\} \right).\end{aligned}$$

The optimal decision rule ϕ^* aims to maximize the later term, which is:

$$\phi^* = \begin{cases} 0, & \text{if } P_0^{\otimes n}\{X^n = x^n\} \geq P_1^{\otimes n}\{X^n = x^n\} \\ 1, & \text{if } P_0^{\otimes n}\{X^n = x^n\} < P_1^{\otimes n}\{X^n = x^n\}. \end{cases}$$

Note that ϕ^* does not choose r, and the problem becomes HW4 problem 6. We have

$$\min_{\phi: \mathcal{X}^n \to \{r,0,1\}} \left\{ \pi_r^{(n)}(\phi) + \pi_e^{(n)}(\phi) \right\} = \min_{\phi: \mathcal{X}^n \to \{0,1\}} \left\{ \pi_r^{(n)}(\phi) + \pi_e^{(n)}(\phi) \right\} = 1 - \mathrm{TV}(P_0, P_1)$$

Grading Policy:

- Any typo: -1

3. (Extremal information measures) [16]

Solution:

a) Note that $\mathsf{Var}[X] = r^2 - \mu^2$. Hence $h(X) \le \frac{1}{2}\log\left(2\pi e\left(r^2-\mu^2\right)\right)$. The equality is achieved when X is a Gaussian random variable with variance $r^2 - \mu^2$, since we also need $\mathsf{E}[X] = \mu$, $X \sim \mathcal{N}(\mu, r^2 - \mu^2)$ is a maximizing distribution.

Grading Policy:

- Write down the maximizing distribution: $+2$
- Reasoning: $+4$
- Any typo: -1

b) Let $X \sim P$, direct calculation gives:

$$-D\left(P||\mathrm{G}(p)\right) = \sum_{x=1}^{\infty} P(x)\log\left(\frac{(1-p)p^{x-1}}{P(x)}\right) = H(X) + \log(1-p) + \log(p)\left(\mu - 1\right).$$

Thus, minimizing $D\left(P||\mathrm{G}(p)\right)$ is equivalent to maximizing $H(X)$ with the constraint $\mathsf{E}[X] = \mu > 1$.
By problem 6 in HW1 we know that $H(X) \le \mu H_{\mathrm{b}}\left(\frac{1}{\mu}\right)$ and $P(x) = \frac{1}{\mu-1}\left(\frac{\mu-1}{\mu}\right)^x = \frac{1}{\mu}\left(1-\frac{1}{\mu}\right)^{x-1}$, which means that $X \sim G(1-\frac{1}{\mu})$.
The minimum value of $D\left(P||\mathrm{G}(p)\right)$ is

$$-\mu H_{\mathrm{b}}\left(\frac{1}{\mu}\right) - \log(1-p) - \log(p)\left(\mu-1\right).$$

Grading Policy:

- Write down $D\left(P||\mathrm{G}(p)\right)$: $+1$
- Find out the correct distribution: $+8$
- Find out the minimum value: $+1$
- Any typo: -1

4. (Sum Channel) [16]

Solution: This is actually HW2 problem 2:

a) By chain rule of mutual information:

$$
\begin{aligned}
&I(X, I; Y) = I(X;Y) + I(I;Y|X) = I(I;Y) + I(X;Y|I) \\
&\Rightarrow I(X;Y) = I(X;Y|I) + I(I;Y) - I(I;Y|X) \\
&\Rightarrow I(X;Y) = I(X;Y|I) + H(I) - H(I|Y) - (H(I|X) - H(I|Y,X)) \\
&\Rightarrow I(X;Y) = I(X;Y|I) + H(I)
\end{aligned}
$$

The last equation is because I is a function of X and also a function of Y (the alphabets are disjoint).

Grading Policy:

- Any typo: -1

b) Note that $P_X = P_{X,I}$ since I is a function of X.

$$
\begin{aligned}
\max_{P_X} I(X;Y) &= \max_{P_{X,I}} (H(I) + I(X;Y|L)) \\
&= \max_{P_I P_{X|I}} (H(I) + I(X;Y|I)) = \max_{P_I} \left(H(I) + \max_{P_{X|I}} I(X;Y|I) \right).
\end{aligned}
$$

The later term can be calculate using the capacity of individual channels, let $p_i = \mathsf{Pr}\{I = i\}$:

$$
\max_{P_{X|I}} I(X;Y|I) = \sum_{i=1}^{l} p_i \max_{P_{X|I=i}} I(X;Y|I=i) = \sum_{i=1}^{l} p_i C^{(i)}.
$$

The equality is achieved when $P_{X|I=i}$ is the capacity achieving distribution for channel $i, \forall i = 1, 2, \cdots, l$.
The problem boils down to solve

$$
C^{\oplus} = \max_{p_1, \cdots, p_l} \sum_{i=1}^{l} \left(p_i \log \frac{1}{p_i} + p_i C^{(i)} \right).
$$

By concavity of $\log(\cdot)$ and Jensen's inequality,

$$
\sum_{i=1}^{l} p_i \log \frac{2^{C^{(i)}}}{p_i} \le \log \sum_{i=1}^{l} p_i \frac{2^{C^{(i)}}}{p_i} = \log \left(\sum_{i=1}^{l} 2^{C^{(i)}} \right).
$$

The equality holds when $2^{C^{(i)}}/p_i = 2^{C^{(j)}}/p_j, \forall i, j \in \{1, \cdots, l\}$.
Which means $p_i = \frac{2^{C^{(i)}}}{\sum_{j=1}^{l} 2^{C^{(j)}}}, \forall i$.
We can conclude that $C^{\oplus} = \log \left(\sum_{i=1}^{l} 2^{C^{(i)}} \right)$.

Grading Policy:

- Write down the maximizing problem : +1
- Find the capacity correctly: +4
- Check the conditions of euqality: +1
- Any typo: −1

c) Denote $P_{X^{(i)}}$ the optimal input distribution for channel i. The optimal input distribution for the sum channel is a mixture distribution:

$$P_X = \sum_{i=1}^{l} \frac{2^{C^{(i)}}}{\sum_{j=1}^{l} 2^{C^{(j)}}} P_{X^{(i)}}.$$

Grading Policy:

- Any typo: −1

5. (Compressing a Uniform Source) [24]

Solution:

a) From lossless source coding theorem we know that $R^* = H(S) = \log(2m)$.

Grading Policy:

- Any typo: -1

b) By HW3 problem 5:

$$R(D) = \begin{cases} \log(2m) - D\log(2m-1) - H_{\mathrm{b}}(D), & 0 \le D \le 1 - \frac{1}{2m}, \\ 0, & D > 1 - \frac{1}{2m}. \end{cases}$$

Grading Policy:

- Write down $\mathrm{D}_{\min}$ and $\mathrm{D}_{\max}$: $+1$ each
- Any typo: -1

c) We know that $D_{\min} = 0$ (choose $\hat{s}$ such that $\hat{s}$ and s has different parity) and $D_{\max} = \frac{1}{2}$ (for arbitrary $\hat{s}$). We want to calculate

$$R(D) = \min_{P_{\hat{S}|S}:\mathsf{E}[d(S,\hat{S})]\le D} I(S;\hat{S}) = \min_{P_{\hat{S}|S}:\mathsf{E}[d(S,\hat{S})]\le D} (H(S) - H(S|\hat{S})).$$

Construct the auxiliary reverse channel $P_{S|\hat{S}}$ and maximize the term $H(S|\hat{S})$.
Denote $\Pr\{\hat{S} = i\}$ by p_i and $\Pr\{S = j|\hat{S} = i\}$ by $q_{i\to j}$. We would like to solve the following problem:

$$\begin{aligned} & \max \sum_{i=1}^{2m} p_i H(q_{i\to 1}, q_{i\to 2}, \cdots, q_{i\to 2m}) \\ \text{subject to:} \quad & \sum_{i=1}^{2m} p_i q_{i\to j} = \Pr\{S = j\} = \frac{1}{2m}, \ \forall j \in \mathcal{S}. \\ & \sum_{k=1}^{m} q_{i\to 2k} \le D \quad \forall i = 2, 4, \cdots, 2m, \\ & \sum_{k=1}^{m} q_{i\to 2k-1} \le D \quad \forall i = 1, 3, \cdots, 2m-1, \\ & \sum_{k=1}^{2m} q_{i\to k} = 1 \quad \forall i \in \mathcal{S}. \end{aligned}$$

By Jensen's inequality:

$$H(q_{i\to 1}, q_{i\to 2}, \cdots, q_{i\to 2m}) \le D\log\left(\frac{m}{D}\right) + (1-D)\log\left(\frac{m}{1-D}\right)$$

$$= \log(m) + H_{\mathrm{b}}(D), \forall i \in \mathcal{S}.$$

The equality holds when

$$\begin{cases} q_{i\to 1} = q_{i\to 3} = \cdots = q_{i\to 2m-1} = \frac{D}{m} & , q_{i\to 2} = q_{i\to 4} = \cdots = q_{i\to 2m} = \frac{1-D}{m}, i \text{ is odd.} \\ q_{i\to 1} = q_{i\to 3} = \cdots = q_{i\to 2m-1} = \frac{1-D}{m} & , q_{i\to 2} = q_{i\to 4} = \cdots = q_{i\to 2m} = \frac{D}{m}, i \text{ is even.} \end{cases}$$

Therefore, $\sum_{i=1}^{2m} p_i H(q_{i\to 1}, q_{i\to 2}, \cdots, q_{i\to 2m}) \leq \log(m) + H_{\mathrm{b}}(D)$. We can choose $p_i = \frac{1}{2m}, \forall i \in \mathcal{S}$ to satisfy the first condition.
We can now conclude that

$$R(D) = \begin{cases} \log(2m) - (\log(m) + H_{\mathrm{b}}(D)) = 1 - H_{\mathrm{b}}(D), & 0 \leq D \leq \frac{1}{2}, \\ 0, & D > \frac{1}{2} \end{cases}.$$

<u>Grading Policy:</u>

- Write down $\mathrm{D}_{\min}$ and $\mathrm{D}_{\max}$: $+1$ each
- Any typo: -1

Alternative solution of c): The problem can be viewed as reconstructing $X = S \pmod 2$ with the Hamming distortion measure. Define the set $\mathcal{X}_0 = \{2, 4, \cdots, 2m\}$ and $\mathcal{X}_1 = \{1, 3, \cdots, 2m-1\}$. Let us build the equivalence:
Call the original lossy source coding problem with alphabet $\mathcal{S}$ and parity distoriton measure P_1 and the lossy source coding problem with alphabet $\{0, 1\}$ and Hamming distortion measure P_2.

- P_1 can reduce to P_2 using the following procedure:
  For message s, compress $x = i$ if $s \in \mathcal{X}_i$, for the reconstruct $\hat{x} = i$ choose a element in $\mathcal{X}_i$ uniformly as $\hat{s}$ as the reconstructed message.
- P_2 can reduce to P_1 using the following procedure:
  For message $x \in \{0, 1\}$, choose arbitrary $s \in \mathcal{X}_x$ and compress it, for the reconstruct $\hat{s}$ let the reconstructed $\hat{x} = i$ if $\hat{s} \in \mathcal{X}_i$.

Moreover, the parity distortion constraint in P_1 is equivalent to the Hamming distortion constraint in P_2, therefore, the problems are equivalent.
Since $X \sim \mathrm{Ber}\left(\frac{1}{2}\right)$, the rate distortion function

$$R(D) = \begin{cases} 1 - H_{\mathrm{b}}(D), & 0 \leq D \leq \frac{1}{2}, \\ 0, & D > \frac{1}{2} \end{cases}.$$