ECE 231A HW 5

Lawrence Liu

December 6, 2022

Problem 1

(a)

First let us prove that

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \max_i \frac{a_i}{b_i}$$

We can prove this through induction, we already have that the base case when $n = 2$ is true. Now we consider the case the case of $n+1$, let us arrange the a_i and b_i in such a way such that $\frac{a_{n+1}}{b_{n+1}}$ be the minimum of the $n+1$ $\frac{a_i}{b_i}$'s. Then we have that

$$\frac{a_{n+1}}{b_{n+1}} \leq \min_{1\leq i\leq n} \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$$

Therefore we have that

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^{n+1} a_i}{\sum_{i=1}^{n+1} b_i} \leq \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \leq \max_i \frac{a_i}{b_i}$$

Thus we have proven that $\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$ is bounded by the minimum and maximum of the $\frac{a_i}{b_i}$'s. Using this property we get that

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)} \leq \max_{x\in\mathcal{X}} \frac{\tilde{P}_X(x) D(W_{Y|X}||P_Y)}{\tilde{P}_X(x)c(x)} = \max_{x\in\mathcal{X}} \frac{D(W_{Y|X}||P_Y)}{c(x)}$$

(b)

We have that

$$\begin{aligned}
\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y) - \sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y) &= \sum_{x\in\mathcal{X}} \tilde{P}_X(x) \left(D(W_{Y|X}||P_Y) - D(W_{Y|X}||P_Y) \right) \\
&= \sum_{x\in\mathcal{X}} \tilde{P}_X(x) \sum_{y\in\mathcal{Y}} W_{Y|X}(y|x) \log\left(\frac{\tilde{P}_Y(y)}{P_Y(y)}\right) \\
&= \sum_{y\in\mathcal{Y}} \log\left(\frac{\tilde{P}_Y(y)}{P_Y(y)}\right) \sum_{x\in\mathcal{X}} \tilde{P}_X(x) W_{Y|X}(y|x) \\
&= \sum_{y\in\mathcal{Y}} \tilde{P}_Y(y) \log\left(\frac{\tilde{P}_Y(y)}{P_Y(y)}\right) \\
&= D_{KL}(\tilde{P}_Y||P_Y) \geq 0
\end{aligned}$$

Therefore we have equality only happens when $\tilde{P}_Y = P_Y$. From this we get that

$$\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y) \geq \sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y)$$

Thus we have that

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||\tilde{P}_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)} \leq \frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x) D(W_{Y|X}||P_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)} \leq \max_{x\in\mathcal{X}} \frac{D(W_{Y|X}||\tilde{P}_Y)}{c(x)}$$

(c)

We have that

$$\sum_{x\in\mathcal{X}} \tilde{P}_X(x)D(W_{Y|X}||P_Y) \leq \sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)\lambda$$

Thus we have that

$$\frac{\sum_{x\in\mathcal{X}} \tilde{P}_X(x)D(W_{Y|X}||P_Y)}{\sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)} \leq \lambda$$

With equality if and only if $\tilde{P}_Y = P_Y$ and $\tilde{P}_X(x) = 0$ for all x where $P_X^*(x) = 0$, because then we will have that

$$\sum_{x\in\mathcal{X}} \tilde{P}_X(x)D(W_{Y|X}||P_Y) = \sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)\lambda$$

(d)

We have that

$$I(X;Y) = \sum_{x\in\mathcal{X}} \tilde{P}_X(x)D(W_{Y|X}||P_Y)$$

and

$$E[c(x)] = \sum_{x\in\mathcal{X}} \tilde{P}_X(x)c(x)$$

Thus we have that from part (c)

$$C_{cost} = \max_{P_X(x)} \frac{I(X;Y)}{E[c(x)]} = \lambda$$

Problem 2

(a)

We have that in order to maximize the differential entropy of Z we want to maximize the variance of Z, this is given by

$$\mathrm{Var}(Z) = E[(Z_1 + Z_2)^2] - (E[Z_1 + Z_2])^2$$

Since Z_1 is independent of Z_2 we have that $E[Z_1 Z_2] = E[Z_1]E[Z_2]$, thus we get that

$$\mathrm{Var}(Z) = E[Z_1^2] + E[Z_2^2] - E^2[Z_1] - E^2[Z_2] \leq 2\sigma^2 - E^2[Z_1] - E^2[Z_2]$$

Therefore in order for this to be maximized, we want $E[Z_1] = E[Z_2] = 0$. Thus we get that both Z_1 and Z_2 are Gaussian with mean 0. And that the maximal differential entropy of Z is given by $\frac{1}{2}\log(4\pi e\sigma^2)$.

(b)

Once again we want to maximize the differential entropy of Z, and thus we want to maximize the variance of Z, this is given by

$$\mathrm{Var}(Z) = E[(\sum_{i=1}^{n} Z_i)^2] - (E[\sum_{i=1}^{n} Z_i])^2$$

We have that

$$\left(\sum_{i=1}^{n} Z_i\right)^2 = \sum_{i=1}^{n} Z_i^2 + \sum_{i\neq j} Z_i Z_j$$

And

$$E^2[\sum_{i=1}^{n} Z_i] = \sum_{i=1}^{n} E^2[Z_i] + \sum_{i\neq j} E[Z_i]E[Z_j]$$

Since Z_i's are independent we have that

$$E[Z_i Z_j] - E[Z_i]E[Z_j] = 0$$

For all $i \neq j$, thus we get that

$$\text{Var}(Z) = \sum_{i=1}^{n} E[Z_i^2] - \sum_{i=1}^{n} E^2[Z_i] = n\sigma^2 - \sum_{i=1}^{n} E^2[Z_i]$$

Therefore we must have that $E[Z_i] = 0$ for all i, thus we get that the maximum differential entropy of Z is given by $\boxed{\frac{1}{2}\log(2\pi e n\sigma^2)}$.

(c)

We have that

$$\text{Var}(Z) = E[(\sum_{i=1}^{n} Z_i)^2] - (E[\sum_{i=1}^{n} Z_i])^2$$

We can rewrite this as

$$\begin{aligned}\text{Var}(Z) &= E[\sum_{i=1}^{n} Z_i^2] - E[\sum_{i=1}^{n} Z_i]^2 + \sum_{i \neq j} E[Z_i Z_j] - E[Z_i]E[Z_j] \\ &= \sum_{i=1}^{n} Var(Z_i) + \sum_{i \neq j} Cov(Z_i, Z_j)\end{aligned}$$

We have that $Cov(Z_i, Z_j) \leq \sqrt{Var(Z_i)Var(Z_j)}$, and since $Var(Z_i) \leq \sigma^2$ we get that

$$Var(z) \leq n^2\sigma^2$$

With equality only happening when $E[Z_i] = 0$, therefore we get that the joint distribution of Z_i's as expressed by a vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ is distrubuted as a multivariate Gaussian with mean $\mu = [0, \ldots, 0]$ and covariance matrix Σ where $\Sigma_{ij} = \sigma^2$ for all $1 \leq i, j \leq n$. And with this the maximum differential entropy of Z is given by $\boxed{\frac{1}{2}\log_2(2\pi e n^2\sigma^2)}$.

Problem 3

(a)

We have:

$$\begin{aligned} I(Y_1, Y_2; X) &= H(Y_1, Y_2) - H(Y_1, Y_2|X) \\ &= H(Y_1, Y_2) - H(Z_1, Z_2) \\ &= H(Y_1, Y_2) - H(Z_1) - H(Z_2) \\ &= H(Y_1, Y_2) - \log(2\pi e\sigma^2) \end{aligned}$$

We have that with $\mathbf{Y} = [Y_1, Y_2]$, and since $E[Y_1^2] = E[Y_2^2] = \sigma^2 + P$ and $E[Y_1Y_2] = P$ we get that $K = \begin{bmatrix} \sigma^2 + P & P \\ P & \sigma^2 + P \end{bmatrix}$ therefore we get that

$$H(Y_1, Y_2) = \frac{1}{2}\log_2(2\pi e\sigma^2 det(K)) = \frac{1}{2}\log_2((2\pi e)^2\sigma^2(\sigma^2 + 2P))$$

Thus we get that

$$I(Y_1, Y_2; X) \leq \frac{1}{2}\log\left(1 + \frac{2P}{\sigma^2}\right)$$

Therefore we get that the channel capacity is $\boxed{\frac{1}{2}\log\left(1 + \frac{2P}{\sigma^2}\right)}$ and that we can achieve this capacity by having an X distributed as a Gaussian with mean 0 and variance P.

(b)

We have that we want to maximize

$$\begin{aligned}
I(Y_1,Y_2,Y_3,Y_4;X) &= H(Y_1,Y_2,Y_3,Y_4) - H(Y_1,Y_2,Y_3,Y_4|X) \\
&= H(Y_1,Y_2,Y_3,Y_4) - H(Z_1,Z_2,Z_3,Z_4) \\
&\le H(Y_1,Y_2) + H(Y_3,Y_4) - H(Z_1) - H(Z_2) - H(Z_3) - H(Z_4) \\
&= H(Y_1,Y_2) + H(Y_3,Y_4) - \log(2\pi e\sigma^2) - \log(2\pi e\sigma^2) \\
&\le \frac{1}{2}\log\left(1+\frac{2P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1+\frac{2P_2}{\sigma_2^2}\right)
\end{aligned}$$

Therefore we have the optimization problem for the channel capacity is just maximizing $\frac{1}{2}\log\left(1+\frac{2P_1}{\sigma_1^2}\right) + \frac{1}{2}\log\left(\frac{2P_2}{\sigma_2^2}\right)$ subject to the constrain that $P_1 + P_2 = P$.

(c)

Using a langrangian multiplier we get that to find the optimal solution we must find P_1 and P_2 that satisfy

$$\nabla f(P_1,P_2) = 0$$

where

$$f(P_1,P_2) = \frac{1}{2}\log\left(\frac{2P_1}{\sigma_1^2}+1\right) + \frac{1}{2}\log\left(\frac{2P_2}{\sigma_2^2}+1\right) + \lambda(P_1+P_2-P)$$

We have that

$$\frac{\partial f}{\partial P_1} = \frac{1}{2P_1+\sigma_1^2} + \lambda = 0$$

and

$$\frac{\partial f}{\partial P_2} = \frac{1}{2P_2+\sigma_2^2} + \lambda = 0$$

Therefore we get that

$$P_1 = -\frac{1}{2\lambda} - \frac{\sigma_1^2}{2}$$

$$P_2 = -\frac{1}{2\lambda} - \frac{\sigma_2^2}{2}$$

Thus we get that

$$\lambda = -\frac{2}{\sigma_1^2 + \sigma_2^2 + 2P}$$

and

$$P_1 = \frac{\sigma_1^2 + \sigma_2^2 + 2P}{4} - \sigma_1^2$$

$$P_2 = \frac{\sigma_1^2 + \sigma_2^2 + 2P}{4} - \sigma_2^2$$

However we need to have that $P_1 \geq 0$ and $P_2 \geq 0$ and therefore we have that

$$P_1 = \boxed{\left(\nu - \frac{\sigma_1^2}{2}\right)^+}$$

$$P_2 = \boxed{\left(\nu - \frac{\sigma_2^2}{2}\right)^+}$$

where ν is chosen such that $(\nu - \frac{\sigma_1^2}{2})^+ + (\nu - \frac{\sigma_2^2}{2})^+ = P$, and $(x)^+ = \max\{0, x\}$. This satisfies the Kuhn-Tucker conditions. Therefore we get that the optimal channel capacity is:

$$\boxed{\log\left(\frac{(2\nu - \sigma_1^2)^+}{\sigma_1^2} + 1\right) + \log\left(\frac{(2\nu - \sigma_2^2)^+}{\sigma_2^2} + 1\right)}$$

(d)

We have that if $P = 3$ and $P_2 = 0$ then $P_1 = 3$ and we get that $\nu = 3 + 1 = 4$ and thus we have that $\sigma_2^2 \geq 2\nu = \boxed{8}$

Problem 4

(a)

We have that we want to minimize

$$I(X;\hat{X}|Y) = H(X|Y) - H(X|Y,\hat{X})$$

When $Y \neq E$, the entropy of $H(X|Y \neq E) = H(X|Y \neq E, \hat{X}) = 0$, thus we get

$$I(X;\hat{X}|Y) = (p)(H(X|Y = E) - H(X|Y = E, \hat{X})) = (p)(H(X) - H(X|\hat{X}))$$

From cover and thomas, we get that the minimum of $H(X) - H(X|\hat{X}) = H(\frac{1}{2}) - H(D_E)$. For some D_E, that is expected distortion given that $Y = E$, we get that if $D > \frac{1}{2}p$ then we can set $D_E = \frac{1}{2}$, and thus we will have that $H(X) - H(X|\hat{X}) = H(\frac{1}{2}) - H(\frac{1}{2}) = 0$. If $D < \frac{1}{2}p$ then we want to "lump" all of the distortion to the case when $Y = E$, therefore we will have $D_E = \frac{D}{p}$ and thus we will have that

$$R_{X|Y}(D) = \boxed{\begin{cases} p\left(1 - H\left(\frac{D}{p}\right)\right) & D < \frac{1}{2}p \\ 0 & D \geq \frac{1}{2}p \end{cases}}$$

(b)

From the previous part we once again have

$$I(X;\hat{X}|Y) = p(H(X) - H(X|\hat{X}))$$

Since the $d(0,1) = d(1,0) = \infty$ we have that if $\hat{X} \neq E$, $\hat{X}$ must be equal to X, Therefore $H(X|\hat{X} \neq E) = 0$. Therefore we get that

$$H(X|\hat{X}) \leq p_E H(X|\hat{X} = E) = DH(X) = p_E H(\frac{1}{2}) = D$$

Where p_E is the probability that $\hat{X} = E$ given that $Y = E$. We have that we want $p_E = 1$, but this is only possible if $D > p$ otherwise we will have $p_E = \frac{D}{p}$. Thus we have that the rate distortion function is

$$R_{X|Y}(D) = \boxed{\begin{cases} p(1 - \frac{D}{p}) & D < p \\ 0 & D \geq p \end{cases}}$$

Problem 5

(a)

We have that $H(X) = \frac{1}{2}\log((2\pi e)^2 2)$ and let $E((X_1 - \underline{X_1})^2) = \sigma_1^2$ and $E((X_2 - \underline{X_2})^2) = \sigma_2^2$ for some $\sigma_1^2 + \sigma_2^2 = \frac{5}{2}$. We have that

$$\begin{aligned} I(X;\underline{X}) &= H(X_1) - H(X|\underline{X}) \\ &= \frac{1}{2}\log((2\pi e)^2 2) - H(X - \underline{X}|\underline{X}) \\ &\leq \frac{1}{2}\log((2\pi e)^2 2) - H(X - \underline{X}) \\ &= \frac{1}{2}\log((2\pi e)^2 2) - H(X_1 - \underline{X_1}) - H(X_2 - \underline{X_2}) \\ &= \frac{1}{2}\log((2\pi e)^2 2) - \frac{1}{2}\log((2\pi e)^2 \sigma_1^2) - \frac{1}{2}\log((2\pi e)^2 \sigma_2^2) \end{aligned}$$

Therefore we have that we want to minimize $\frac{1}{2}\log\left(\frac{4}{\sigma_1^2\sigma_2^2}\right)$ subject to the constraint that $\sigma_1^2 + \sigma_2^2 = \frac{5}{2}$.

(b)

To solve for σ_1^2 and σ_2^2 we use a langrangian multiplier and get that we want to find the σ_1^2 and σ_2^2 that satisfy

$$\nabla f(\sigma_1^2, \sigma_2^2) = 0$$

where

$$f(\sigma_1^2, \sigma_2^2) = \frac{1}{2}\log\left(\frac{2}{\sigma_1^2\sigma_2^2}\right) + \lambda(\sigma_1^2 + \sigma_2^2 - \frac{5}{2})$$

We have that

$$\frac{\partial f}{\partial \sigma_1^2} = \lambda - \frac{1}{2\sigma_1^2} = 0$$

and

$$\frac{\partial f}{\partial \sigma_2^2} = \lambda - \frac{1}{2\sigma_2^2} = 0$$

Therefore we get that

$$\sigma_1^2 = \frac{1}{2\lambda}$$

and

$$\sigma_2^2 = \frac{1}{2\lambda}$$

Thus we get that

$$\lambda = \frac{1}{\frac{5}{2}} = \frac{2}{5}$$

and thus we get that

$$\sigma_1^2 = \boxed{\frac{5}{4}}$$

$$\sigma_2^2 = \boxed{\frac{5}{4}}$$

And thus we will send X_1 through a gaussian channel with $Z_1 \sim \mathcal{N}(0, \frac{5}{4})$ and X_2 through a gaussian channel with $Z_2 \sim \mathcal{N}(0, \frac{5}{4})$.

(c)

We have that we can transform Y_1 abd Y_2 into

$$Y_1' = Y_1 + Y_2$$

$$Y_2' = Y_1 - Y_2$$

These are independent, and both are gaussian with mean 0 and variance of 4 and 2 respectively. So we can send Y_1' and Y_2' through gaussian channels with $Z_1' \sim \mathcal{N}(0, \sigma_1^2)$ and $Z_2' \sim \mathcal{N}(0, \sigma_2^2)$ respectively.with outputs of $\hat{Y}_1'$ and $\hat{Y}_2'$, then we will have that the rate disortion measure is:

$$\begin{aligned} D &= E\left[\left(\frac{Y_1' + Y_2'}{2} - \frac{\hat{Y}_1' + \hat{Y}_2'}{2}\right)^2\right] + E\left[\left(\frac{Y_1' - Y_2'}{2} - \frac{\hat{Y}_1' - \hat{Y}_2'}{2}\right)^2\right] \\ &= \frac{\sigma_1^2}{2} + \frac{\sigma_2^2}{2} \end{aligned}$$

Thus we have that in order to maximize the rate disortion function, we must maximize

$$\frac{1}{2}\log((2\pi e)^2 8) - \frac{1}{2}\log((2\pi e)^2\sigma_1^2) - \frac{1}{2}\log((2\pi e)^2\sigma_2^2)$$

With the constraint that $\sigma_1^2 + \sigma_2^2 = 5$. We have that using a langrangian multiplier we get that we want to find the σ_1^2 and σ_2^2 that satisfy

$$\nabla f(\sigma_1^2, \sigma_2^2) = 0$$

where

$$f(\sigma_1^2, \sigma_2^2) = \frac{1}{2}\log\left(\frac{8}{\sigma_1^2\sigma_2^2}\right) + \lambda(\sigma_1^2 + \sigma_2^2 - 5)$$

Once again we get that

$$\frac{\partial f}{\partial \sigma_1^2} = \lambda - \frac{1}{2\sigma_1^2} = 0$$

and

$$\frac{\partial f}{\partial \sigma_2^2} = \lambda - \frac{1}{2\sigma_2^2} = 0$$

Therefore we get that

$$\sigma_1^2 = \frac{1}{2\lambda}$$

and

$$\sigma_2^2 = \frac{1}{2\lambda}$$

Thus we get that

$$\lambda = \frac{1}{5}$$

and thus we get that

$$\sigma_1^2 = \boxed{\frac{5}{2}}$$

$$\sigma_2^2 = \boxed{\frac{5}{2}}$$