EE276 Homework #6

Due on March 1, 5pm

1. **Rate distortion for uniform source with Hamming distortion.**
Consider a source X uniformly distributed on the set $\{1, 2, ..., m\}$. Find the rate distortion function for this source with Hamming distortion, i.e.,

$$d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x} \\ 1, & x \neq \hat{x} \end{cases}$$

via the following steps:

   (a) Argue that $R(D) = 0$ when $D \geq 1 - \frac{1}{m}$.

   (b) Show that for $D \leq 1 - \frac{1}{m}$, $I(X; \hat{X}) \geq \log_2 m - h_2(D) - D \log_2(m-1)$ for any joint distribution $(X, \hat{X})$ satisfying the distortion constraint D.
   Hint: Fano's inequality.

   (c) Find distribution $p(\hat{x}|x)$ that achieves the above lower bound when $0 \leq D \leq 1 - \frac{1}{m}$.

   (d) Use the above parts to write down the rate-distortion function $R(D)$ for $D \geq 0$.

Solution: **Rate distortion for uniform source with Hamming distortion.**

X is uniformly distributed on the set $\{1, 2, ..., m\}$. The distortion measure is

$$d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x} \\ 1, & x \neq \hat{x} \end{cases}$$

For (a), it's enough to see that setting $\hat{X} = 1$ independently of X achieves distortion $1 - 1/m$.

For (b), consider any joint distribution that satisfies the distortion constraint D. Since $D = Pr(X \neq \hat{X})$, we have by Fano's inequality

$$H(X|\hat{X}) \leq h_2(D) + D \log(m-1)$$

and hence

$$\begin{aligned} I(X; \hat{X}) &= H(X) - H(X|\hat{X}) \\ &\geq \log m - h_2(D) - D \log(m-1) \end{aligned}$$

For (c), we can achieve this lower bound by choosing $p(\hat{x})$ to be the uniform distribution, and the conditional distribution of $p(x|\hat{x})$ to be

$$p(\hat{x}|X) = \begin{cases} 1 - D, & x = \hat{x} \\ \frac{D}{m-1}, & x \neq \hat{x} \end{cases}$$

It is easy to verify that this gives the right distribution on X and satisfies the bound with equality for $D < 1 - 1/m$.

For (d),

$$R(D) = \begin{cases} \log m - h_2(D) - D\log(m-1), & 0 \le D \le 1 - \frac{1}{m} \\ 0, & D > 1 - \frac{1}{m} \end{cases}$$

2. **Convexity of rate distortion function.**
Assume $(X, Y) \sim p(x, y) = p(x)p(y|x)$. In this problem, you will show that for fixed $p(x)$, $I(X;Y)$ is a convex function of $p(y|x)$.

   (a) The log sum inequality states that for n positive numbers $a_1, a_2, \cdots, a_n$, and $b_1, b_2, \cdots, b_n$, we have

   $$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \ge \left(\sum_{i=1}^{n} a_i\right) \log\left(\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}\right)$$

   with equality if and only if $\frac{a_i}{b_i}$ =constant. Using this inequality (you don't have to prove this inequality), show that $D(p||q)$ is convex in (p, q), i.e.,

   $$\lambda D(p_1||q_1) + (1-\lambda)D(p_2||q_2) \ge D(\lambda p_1 + (1-\lambda)p_2||\lambda q_1 + (1-\lambda)q_2)$$

   (b) Let $p_1(y|x)$ and $p_2(y|x)$ be two different conditional distributions. For $i \in \{1, 2\}$, let $p_i(x, y) = p_i(y|x)p(x)$, i.e., their corresponding joint distributions. For $0 \le \lambda \le 1$, let $p_\lambda(y|x) \triangleq \lambda p_1(y|x) + (1-\lambda)p_2(y|x)$. Show that

   $$p_\lambda(y) = \lambda p_1(y) + (1-\lambda)p_2(y)$$

   (c) The mutual information between random variables X and Y can be alternatively written as

   $$I(X;Y) = D(p(x, y)||p(x)p(y))$$

   Using this in addition to the results of the previous parts show that for fixed $p(x)$, $I(X;Y)$ is convex in $p(y|x)$.

   (d) Using the previous part, show that the rate distortion function $R^{(I)}(D)$ is convex in the distortion parameter D.

Solution: **Solution: Convexity of rate distortion function.**

(a) By definition,

$$
\begin{aligned}
D(\lambda p_1 + (1-\lambda)p_2 || \lambda q_1 + (1-\lambda)q_2) \\
&= \sum_{x\in\mathcal{X}} (\lambda p_1(x) + (1-\lambda)p_2(x)) \log \frac{\lambda p_1(x) + (1-\lambda)p_2(x)}{\lambda q_1(x) + (1-\lambda)q_2(x)} \\
&\overset{(a)}{\leq} \sum_{x\in\mathcal{X}} \lambda p_1(x) \log \frac{\lambda p_1(x)}{\lambda q_1(x)} + (1-\lambda)p_2(x) \log \frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)} \\
&= \sum_{x\in\mathcal{X}} \lambda p_1(x) \log \frac{p_1(x)}{q_1(x)} + (1-\lambda)p_2(x) \log \frac{p_2(x)}{q_2(x)} \\
&= \lambda D(p_1||q_1) + (1-\lambda)D(p_2||q_2)
\end{aligned}
$$

where (a) is because of the log-sum inequality.

(b) We have

$$
\begin{aligned}
p_\lambda(y) &= \sum_{x\in\mathcal{X}} p_\lambda(y|x)p(x) \\
&= \sum_{x\in\mathcal{X}} (\lambda p_1(y|x) + (1-\lambda)p_2(y|x))p(x) \\
&= \lambda \sum_{x\in\mathcal{X}} p_1(y|x)p(x) + (1-\lambda) \sum_{x\in\mathcal{X}} p_2(y|x)p(x) \\
&= \lambda p_1(y) + (1-\lambda)p_2(y).
\end{aligned}
$$

(c) Let $p_1(y|x)$ and $p_2(y|x)$ be two different conditional distributions, and let $I_1(X;Y)$ and $I_2(X;Y)$ denote that respective mutual information between X and Y when $p(x)$ is fixed. Note that

$$
\begin{aligned}
\lambda I_1(X;Y) + (1-\lambda)I_2(X;Y) \\
&= \lambda D(p_1(x,y)||p(x)p_1(y)) + (1-\lambda)D(p_2(x,y)||p(x)p_2(y)) \\
&\geq D(\lambda p_1(x,y) + (1-\lambda)p_2(x,y)||\lambda p(x)p_1(y) + (1-\lambda)p(x)p_2(y)) \\
&= D(p_\lambda(x,y)||p(x)p_\lambda(y)) \\
&= I_\lambda(X;Y).
\end{aligned}
$$

where $I_l(X;Y)$ corresponds to the mutual information between X and Y when the conditional distribution of Y given X is $p_\lambda(y|x)$.

(d) Consider distortions D_1 and D_2. We need to show that

$$
R^{(I)}(\lambda D_1 + (1-\lambda)D_2) \leq \lambda R^{(I)}(D_1) + (1-\lambda)R^{(I)}(D_2)
$$

for any $\lambda \in [0,1]$. To show this, consider the joint distributions achieving the rate-distortion optimum at D_1 and D_2, $p_1(x,\hat{x}) = p(x)p_1(\hat{x}|x)$ and $p_2(x,\hat{x}) = p(x)p_2(\hat{x}|x)$. Also consider the distribution $p_\lambda = \lambda p_1 + (1-\lambda)p_2$. Since distortion is a linear function of the joint probability distribution, the distortion for p_λ is at most $\lambda D_1 + (1-\lambda)D_2$. By definition of $R^{(I)}(D)$,

$$\begin{aligned} R^{(I)}(\lambda D_1 + (1-\lambda)D_2) &\le I_\lambda(X;\hat{X}) \\ &\le \lambda I_1(X;\hat{X}) + (1-\lambda)I_2(X;\hat{X}) \\ &= \lambda R^{(I)}(D_1) + (1-\lambda)R^{(I)}(D_2) \end{aligned}$$

where I_λ, I_1 and I_2 denote the mutual informations when the distribution is p_λ, p_1 and p_2, respectively. The second inequality uses the convexity of mutual information proved in part (c).

3. **Shannon lower bound.**
Let X be a continuous random variable with mean zero and variance σ^2. $R(D)$ is the corresponding rate-distortion function for mean-squared distortion.

(a) Show the lower bound:

$$h(X) - \frac{1}{2}\log(2\pi e D) \le R(D).$$

(b) Using the joint distribution shown in Figure 1, show the upper bound on $R(D)$:

$$R(D) \le \frac{1}{2}\log\frac{\sigma^2}{D} \tag{1}$$

Are Gaussian random variables harder or easier to describe than other random variables with the same variance?

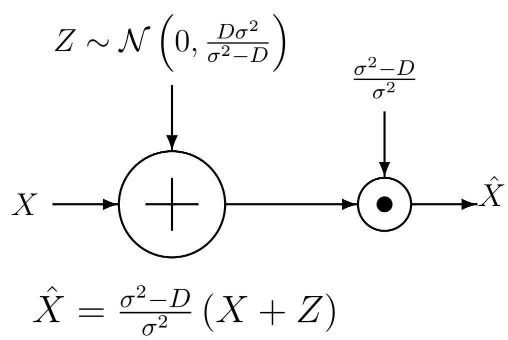

Figure 1: Joint distribution for upper bound on rate distortion function. The circle with the dot represents multiplication.

Solution: **Solution: Shannon Lower Bound.**

(a) To prove the lower bound, we use the same techniques as used for the Guassian rate distortion function. Let $(X, \hat{X})$ be random variables such that $\mathcal{E}(X - \hat{X})^2 \leq D$. Then

$$\begin{aligned}
I(X;\hat{X}) &= h(X) - h(X|\hat{X}) & (2)\\
&= h(X) - h(X - \hat{X}|\hat{X}) & (3)\\
&\geq h(X) - h(X - \hat{X}) & (4)\\
&\geq h(X) - h(\mathcal{N}(0, \mathcal{E}(X - \hat{X})^2)) & (5)\\
&= h(X) - \frac{1}{2}\log(2\pi e)\mathcal{E}(X - \hat{X})^2 & (6)\\
&\geq h(X) - \frac{1}{2}\log(2\pi e)D. & (7)
\end{aligned}$$

(b) Note that you could have used any constant instead of $(\sigma^2 - D)/\sigma^2$, since the mutual information in question will be the same for all choices. Here, we'll use the one in the problem.

To prove the upper bound, we consider the joint distribution as shown in Figure 1, and calculate the distortion and the mutual information between X and $\hat{X}$. Since

$$\hat{X} = \frac{\sigma^2 - D}{\sigma^2}\left(X + Z\right), \tag{8}$$

we have

$$\begin{aligned}
\mathcal{E}(X - \hat{X})^2 &= \mathcal{E}\left(\frac{D}{\sigma^2}X - \frac{\sigma^2 - D}{\sigma^2}Z\right)^2 & (9)\\
&= \left(\frac{D}{\sigma^2}\right)^2 \mathcal{E}X^2 + \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \mathcal{E}Z^2 & (10)\\
&= \left(\frac{D}{\sigma^2}\right)^2 \sigma^2 + \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \frac{D\sigma^2}{\sigma^2 - D} & (11)\\
&= D, & (12)
\end{aligned}$$

since X and Z are independent and zero mean. Also the mutual information is

$$\begin{aligned}
I(X;\hat{X}) &= h(\hat{X}) - h(\hat{X}|X) & (13)\\
&= h(\hat{X}) - h(\frac{\sigma^2 - D}{\sigma^2}Z). & (14)
\end{aligned}$$

Now

$$\begin{aligned}
\mathcal{E}\hat{X}^2 &= \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \mathcal{E}(X + Z)^2 & (15)\\
&= \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 (\mathcal{E}X^2 + \mathcal{E}Z^2) & (16)\\
&= \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 \left(\sigma^2 + \frac{D\sigma^2}{\sigma^2 - D}\right) & (17)\\
&= \sigma^2 - D. & (18)
\end{aligned}$$

Hence, we have

$$\begin{aligned}
I(X;\hat{X}) &= h(\hat{X}) - h(\frac{\sigma^2 - D}{\sigma^2}Z) && (19)\\
&= h(\hat{X}) - h(Z) - \log\frac{\sigma^2 - D}{\sigma^2} && (20)\\
&\leq h(\mathcal{N}(0,\sigma^2 - D)) - \frac{1}{2}\log(2\pi e)\frac{D\sigma^2}{\sigma^2 - D} - \log\frac{\sigma^2 - D}{\sigma^2} && (21)\\
&= \frac{1}{2}\log(2\pi e)(\sigma^2 - D) - \frac{1}{2}\log(2\pi e)\frac{D\sigma^2}{\sigma^2 - D} - \frac{1}{2}\log\left(\frac{\sigma^2 - D}{\sigma^2}\right)^2 && (22)\\
&= \frac{1}{2}\log\frac{\sigma^2}{D}, && (23)
\end{aligned}$$

which combined with the definition of the rate distortion function gives us the required upper bound.

For a Gaussian random variable, $h(X) = \frac{1}{2}\log(2\pi e)\sigma^2$ and the lower bound is equal to the upper bound. For any other random variable, the lower bound is strictly less than the upper bound and hence non-Gaussian random variables cannot require more bits to describe to the same accuracy than the corresponding Gaussian random variables. This is not surprising, since the Gaussian random variable has the maximum entropy and we would expect that it would be the most difficult to describe.

4. **Rate distortion for two independent sources.** Let $\{X_i\}$ be iid $\sim p(x)$ with distortion $d(x,\hat{x})$ and rate distortion function $R_X(D)$. Similarly, let $\{Y_i\}$ be iid $\sim p(y)$ with distortion $d(y,\hat{y})$ and rate distortion function $R_Y(D)$.

   Suppose the $\{X_i\}$ process and the $\{Y_i\}$ process are independent of each other.

   Suppose we now wish to describe the process $\{(X_i,Y_i)\}$ subject to distortions $\mathcal{E}[d(X,\hat{X})] \leq D_1$ and $\mathcal{E}[d(Y,\hat{Y})] \leq D_2$. Thus a rate $R_{X,Y}(D_1,D_2)$ is sufficient, where

$$R_{X,Y}(D_1,D_2) = \min_{p(\hat{x},\hat{y}|x,y):\mathcal{E}[d(X,\hat{X})]\leq D_1,\mathcal{E}[d(Y,\hat{Y})]\leq D_2} I(X,Y;\hat{X},\hat{Y})$$

   Express $R_{X,Y}(D_1,D_2)$ in terms of $R_X(D_1)$ and $R_Y(D_2)$. Can one simultaneously compress two independent sources better than by compressing the sources individually?

   Solution: *Rate distortion for two independent sources*

   (a) Given that X and Y are independent, we have

$$p(x,y,\hat{x},\hat{y}) = p(x)p(y)p(\hat{x},\hat{y}|x,y) \quad (24)$$

   Then

$$\begin{aligned}
I(X,Y;\hat{X},\hat{Y}) &= H(X,Y) - H(X,Y|\hat{X},\hat{Y}) && (25)\\
&= H(X) + H(Y) - H(X|\hat{X},\hat{Y}) - H(Y|X,\hat{X},\hat{Y}) && (26)\\
&\geq H(X) + H(Y) - H(X|\hat{X}) - H(Y|\hat{Y}) && (27)\\
&= I(X;\hat{X}) + I(Y;\hat{Y}) && (28)
\end{aligned}$$

where the inequality follows from the fact that conditioning reduces entropy. Therefore

$$\begin{aligned}
R_{X,Y}(D_1, D_2) &= \min_{p(\hat{x},\hat{y}|x,y): Ed(X,\hat{X}) \le D_1, Ed(Y,\hat{Y}) \le D_2} I(X, Y; \hat{X}, \hat{Y}) && (29)\\
&\ge \min_{p(\hat{x},\hat{y}|x,y): Ed(X,\hat{X}) \le D_1, Ed(Y,\hat{Y}) \le D_2} \Big(I(X; \hat{X}) + I(Y; \hat{Y})\Big) && (30)\\
&= \min_{p(\hat{x}|x): Ed(X,\hat{X}) \le D_1} I(X; \hat{X}) + \min_{p(\hat{y}|y): Ed(Y,\hat{Y}) \le D_2} I(Y; \hat{Y}) && (31)\\
&= R_X(D_1) + R_Y(D_2) && (32)
\end{aligned}$$

(b) If

$$p(x, y, \hat{x}, \hat{y}) = p(x)p(y)p(\hat{x}|x)p(\hat{y}|y), \tag{33}$$

then

$$\begin{aligned}
I(X, Y; \hat{X}, \hat{Y}) &= H(X, Y) - H(X, Y|\hat{X}, \hat{Y}) && (34)\\
&= H(X) + H(Y) - H(X|\hat{X}, \hat{Y}) - H(Y|X, \hat{X}, \hat{Y}) && (35)\\
&= H(X) + H(Y) - H(X|\hat{X}) - H(Y|\hat{Y}) && (36)\\
&= I(X; \hat{X}) + I(Y; \hat{Y}) && (37)
\end{aligned}$$

Let $p(x, \hat{x})$ be a distribution that achieves the rate distortion $R_X(D_1)$ at distortion D_1 and let $p(y, \hat{y})$ be a distribution that achieves the rate distortion $R_Y(D_2)$ at distortion D_2. Then for the product distribution $p(x, y, \hat{x}, \hat{y}) = p(x, \hat{x})p(y, \hat{y})$, where the component distributions achieve rates $(D_1, R_X(D_1))$ and $(D_2, R_X(D_2))$, the mutual information corresponding to the product distribution is $R_X(D_1) + R_Y(D_2)$. Thus

$$R_{X,Y}(D_1, D_2) = \min_{p(\hat{x},\hat{y}|x,y): Ed(X,\hat{X}) \le D_1, Ed(Y,\hat{Y}) \le D_2} I(X, Y; \hat{X}, \hat{Y}) = R_X(D_1) + R_Y(D_2) \tag{38}$$

Thus by using the product distribution, we can achieve the sum of the rates.

Therefore the total rate at which we encode two independent sources together with distortions D_1 and D_2 is the same as if we encoded each of them separately.

5. **Distortion-rate function.** Let

$$D(R) = \min_{p(\hat{x}|x): I(X;\hat{X}) \le R} \mathcal{E}[d(X, \hat{X})] \tag{39}$$

be the distortion rate function.

(a) Is $D(R)$ increasing or decreasing in R?

(b) Is $D(R)$ convex or concave in R?

(c) Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim p(x)$. Suppose one is given a code $(X^n, \hat{X}^n)$ with

$$\frac{1}{n} I(X^n; \widehat{X}^n) \le R$$

and resulting distortion $D = \mathcal{E}[d(X^n, \hat{X}^n)]$. We want to show that $D \geq D(R)$. Give reasons for the following steps in the proof:

$$\begin{aligned}
D &= \mathcal{E}[d(X^n, \hat{X}^n(i(X^n)))] && (40)\\
&\stackrel{(a)}{=} \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right] && (41)\\
&\stackrel{(b)}{=} \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}[d(X_i, \hat{X}_i)] && (42)\\
&\stackrel{(c)}{\geq} \frac{1}{n}\sum_{i=1}^{n} D\left(I(X_i; \hat{X}_i)\right) && (43)\\
&\stackrel{(d)}{\geq} D\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i; \hat{X}_i)\right) && (44)\\
&\stackrel{(e)}{\geq} D\left(\frac{1}{n} I(X^n; \hat{X}^n)\right) && (45)\\
&\stackrel{(f)}{\geq} D(R) && (46)
\end{aligned}$$

Solution: *Distortion rate function.*

(a) Since for larger values of R, the minimization in

$$D(R) = \min_{p(\hat{x}|x): I(X;\hat{X}) \leq R} Ed(X, \hat{X}) \qquad (47)$$

is over a larger set of possible distributions, the minimum has to be at least as small as the minimum over the smaller set. Thus $D(R)$ is a nonincreasing function of R.

(b) By similar arguments as in Lemma 10.4.1, we can show that $D(R)$ is a convex function of R. Consider two rate distortion pairs (R_1, D_1) and (R_2, D_2) which lie on the distortion-rate curve. Let the joint distributions that achieve these pairs be $p_1(x, \hat{x}) = p(x)p_1(\hat{x}|x)$ and $p_2(x, \hat{x}) = p(x)p_2(\hat{x}|x)$. Consider the distribution $p_{l = lp_1 + (1-l)p_2}$. Since the distortion is a linear function of the distribution, we have $D(p_l) = lD_1 + (1-l)D_2$. Mutual information, on the other hand, is a convex function of the conditional distribution (Theorem 2.7.4) and hence

$$I_{p_l}(X; \hat{X}) \leq l I_{p_1}(X; \hat{X}) + (1-l) I_{p_2}(X; \hat{X}) = lR_1 + (1-l)R_2$$

Therefore we can achieve a distortion $lD_1 + (1-l)D_2$ with a rate less than $lR_1 + (1-l)R_2$ and hence

$$\begin{aligned}
D(R_{l)} &\leq D_{p_l(X;\hat{X})} && (48)\\
&= lD(R_1) + (1-l)D(R_2), && (49)
\end{aligned}$$

which proves that $D(R)$ is a convex function of R.

6.

$$
\begin{aligned}
D &= Ed(X^n, \hat{X}^n(i(X^n))) && (50)\\
&\overset{(a)}{=} E\frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i) && (51)\\
&\overset{(b)}{=} \frac{1}{n}\sum_{i=1}^{n} Ed(X_i, \hat{X}_i) && (52)\\
&\overset{(c)}{\geq} \frac{1}{n}\sum_{i=1}^{n} D\left(I(X_i; \hat{X}_i)\right) && (53)\\
&\overset{(d)}{\geq} D\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i; \hat{X}_i)\right) && (54)\\
&\overset{(e)}{\geq} D\left(\frac{1}{n} I(X^n; \hat{X}^n)\right) && (55)\\
&\overset{(f)}{\geq} D(R) && (56)
\end{aligned}
$$

(a) follows from the definition of distortion for sequences
(b) from exchanging summation and expectation
(c) from the definition of the distortion rate function based on the joint distribution $p(x_i, \hat{x}_i)$,
(d) from Jensen's inequality and the convexity of $D(R)$
(e) from the fact that

$$
\begin{aligned}
I(X^n; \hat{X}^n) &= H(X^n) - H(X^n|\hat{X}^n) && (57)\\
&= \sum_{i=1}^{n} H(X_i) - H(X^n|\hat{X}^n) && (58)\\
&= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}^n, X_{i-1}, \ldots, X_1) && (59)\\
&\geq \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}_i) && (60)\\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_i) && (61)
\end{aligned}
$$

and
(f) follows from the definition of the distortion rate function, since $\frac{1}{n} I(X^n; \widehat{X}^n) \leq R$