EE276: Homework #8

Due on Friday March 13, 11:59pm - Gradescope entry code: E6VP4X

1. **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:

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

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

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

   Are Gaussian random variables harder or easier to describe in bits - in the sense of achieving small mean squared error distortion - than other random variables with the same variance?

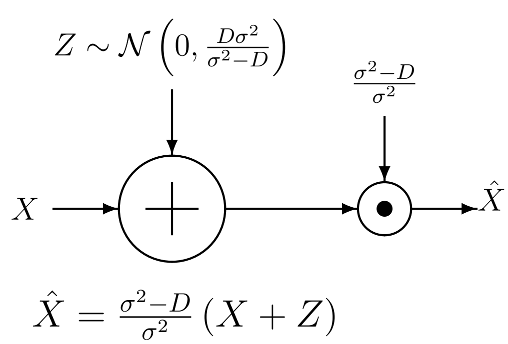

Figure 1: Joint distribution for upper bound on rate distortion function. The circle with the dot represents multiplication.

2. **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 $0 \leq D \leq 1-\frac{1}{m}$.
Hint: Consider the form below.

$$p(\hat{x}|x) = \begin{cases} a, & x = \hat{x} \\ b, & x \neq \hat{x} \end{cases}$$

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

3. **Rate distortion for two independent sources.** Can one simultaneously compress two independent sources better than by compressing the sources individually? This problem addresses this question. Let $\{X_i\}$ be iid $\sim p(x)$ with distortion $d_1(x, \hat{x})$ and rate distortion function $R_X(D)$. Similarly, let $\{Y_i\}$ be iid $\sim p(y)$ with distortion $d_2(y, \hat{y})$ 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 $\mathbb{E}[d_1(X,\hat{X})] \leq D_1$ and $\mathbb{E}[d_2(Y,\hat{Y})] \leq D_2$. Thus a rate $R_{X,Y}(D_1, D_2)$ is sufficient, where

$$R_{X,Y}(D_1, D_2) = \min_{p(\hat{x},\hat{y}|x,y):\mathbb{E}[d_1(X,\hat{X})]\leq D_1, \mathbb{E}[d_2(Y,\hat{Y})]\leq D_2} I(X,Y;\hat{X},\hat{Y})$$

   (a) Show that
$$I(X,Y;\hat{X},\hat{Y}) \geq I(X;\hat{X}) + I(Y;\hat{Y})$$

   (b) Hence conclude
$$R_{X,Y}(D_1, D_2) \geq R_X(D_1) + R_Y(D_2).$$

   (c) Show that the above actually holds with equality.

4. In what follows, all random variables have finite alphabet, and all pmfs are defined on finite alphabets. First, recall the definitions from class:

   Strongly typical. A sequence $x^n \in \mathcal{X}^n$ is *strongly δ-typical* with respect to a probability mass function $P \in \mathcal{M}(\mathcal{X})$ if

$$|P_{x^n}(a) - P(a)| \leq \delta \cdot P(a), \quad \forall a \in \mathcal{X}$$

   where $P_{x^n}(a)$ is the empirical probability of seeing a based on sequence x^n.

   In words, a sequence is strongly δ-typical with respect to P if its empirical distribution is close to the probability mass function P. (δ is some fixed number, typically small.) The *strongly δ-typical set* (ie. strongly typical set) of P, $T_\delta(P)$, is defined as the set of all sequences that are strongly δ-typical with respect to P, i.e.

$$T_\delta(P) = \{x^n : |P_{x^n}(a) - P(a)| \leq \delta \cdot P(a), \forall a \in \mathcal{X}\}$$

Recall: the *weakly ϵ-typical set*, which you are familiar with, of an IID source P is defined as

$$A_\epsilon(P) := \{x^n : \left| -\frac{1}{n} \log P(x^n) - H(P) \right| \leq \epsilon\}.$$

(Strongly) Jointly typical. In the following, we refer to the sequences $x^n = (x_1, x_2, \ldots, x_n)$, $x_i \in \mathcal{X}$ and $y^n = (y_1, y_2, \ldots, y_n)$, $y_i \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets.
The *joint empirical distribution* of (x^n, y^n) is:

$$P_{x^n,y^n}(x, y) = \frac{1}{n} N(x, y|x^n, y^n)$$

where

$$N(x, y|x^n, y^n) := \sum_{i=1}^{n} \mathbf{1}_{\{x_i=x, y_i=y\}}.$$

(x^n, y^n) is *jointly δ-typical* with respect to $P \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ if

$$|P_{x^n,y^n}(x, y) - P(x, y)| \leq \delta \cdot P(x, y), \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}.$$

The *jointly δ-typical set* with respect to $P \in \mathcal{M}(\mathcal{X} \times \mathcal{Y})$ is

$$T_\delta(P) = \{(x^n, y^n) : (x^n, y^n) \text{ is jointly } \delta\text{-typical with respect to } P\}.$$

(a) Let Z be a random variable with pmf p_z and alphabet $\mathcal{Z}$. Let $T_\delta(p_z)$ be the set of strongly δ-typical sequences with respect to p_z.
Show that for any nonnegative $g : \mathcal{Z} \to \mathbb{R}^+$, and $z^n \in T_\delta(p_z)$, we have

$$\left| \frac{1}{n} \sum_{i=1}^{n} g(z_i) - \mathbb{E}[g(Z)] \right| \leq \delta \mathbb{E}[g(Z)].$$

In what follows, it may be useful to invoke part (a) in your arguments.

(b) Let (X, Y) be random variables with joint pmf $p_{x,y}$. Let $T_\delta(p_{x,y})$ be the set of jointly δ-typical sequences with respect to $p_{x,y}$. Show that

$$\frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) \leq (1 + \delta) \mathbb{E}[d(X, Y)]$$

for any distortion function $d(x, y)$ and $(x^n, y^n) \in T_\delta(p_{x,y})$.

(c) Let $A_\epsilon(p)$ denote the (weakly) typical set with respect to p, and $T_\delta(p)$ be the set of δ-typical sequences with respect to p. Show that

$$T_\delta(p) \subseteq A_\epsilon(p)$$

for $\epsilon = \delta H(p)$.

(d) Let Q and P be pmfs over $\mathcal{X}$, and $T_\delta(P)$ be the set of δ-typical sequences with respect to P. Assuming $D(P\|Q)$ is finite, show that

$$Q(T_\delta(P)) \doteq 2^{-n(D(P\|Q) - \alpha(\delta))},$$

where $\alpha(\delta) \geq 0$ and $\alpha(\delta) \to 0$ as $\delta \to 0$.