Homework Set no. 10

1. **Geometric interpretation of channel capacity.** Consider a channel $\{\mathcal{X}, p(y|x), \mathcal{Y}\}$, and assume that the channel input alphabet and output alphabet are of size m, e.g., let $\mathcal{X} = \mathcal{Y} = \{1, 2, \ldots, m\}$. We can write the probability transition function of the channel as a matrix:

$$p(y|x) = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{bmatrix} \tag{1}$$

where $\mathbf{r}_i$ is the i-th row of the probability transistion matrix, i.e, $\mathbf{r}_i = (p(y = 1|x = i), p(y = 2|x = i), \ldots, p(y = m|x = i))$. Consider an input distribution $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ for the channel, and let $\mathbf{q} = (q_1, q_2, \ldots, q_m)$ be the corresponding output distribution, i.e., $\mathbf{q} = \sum_i p_i \mathbf{r}_i$.

(a) Show that for this channel

$$\sum_i p_i D(\mathbf{r}_i || \mathbf{q}) \le C \tag{2}$$

(b) Let $\mathbf{q}'$ be another distribution on $\mathcal{Y}$. Show that

$$\sum_i p_i D(\mathbf{r}_i || \mathbf{q}) \le \sum_i p_i D(\mathbf{r}_i || \mathbf{q}') \tag{3}$$

Use this to argue

$$C = \max_{\mathbf{p}} \sum_i p_i D(\mathbf{r}_i || \mathbf{q}) = \max_{\mathbf{p}} \min_{\mathbf{q}'} \sum_i p_i D(\mathbf{r}_i || \mathbf{q}') \tag{4}$$

and that for any distribution $\mathbf{q}'$,

$$C \le \max_i D(\mathbf{r}_i || \mathbf{q}'). \tag{5}$$

(c) Combining the first two parts, we have

$$\sum_i p_i D(\mathbf{r}_i || \mathbf{q}) \le C \le \min_{\mathbf{q}'} \max_i D(\mathbf{r}_i || \mathbf{q}') \tag{6}$$

Now justify the following argument:

$$\max_{\mathbf{p}} \min_{\mathbf{q}'} \sum_i p_i D(\mathbf{r}_i || \mathbf{q}') \quad \ge \quad \min_{\mathbf{q}'} \max_i D(\mathbf{r}_i || \mathbf{q}') \tag{7}$$

(This is a very general result that max min $\ge$ min max, and can be argued from first principles.)

Combining the previous two equations, we have

$$\max_{\mathbf{p}} \sum_i p_i D(\mathbf{r}_i || \mathbf{q}) = C = \min_{\mathbf{q}'} \max_i D(\mathbf{r}_i || \mathbf{q}') \tag{8}$$

A simple geometric interpretaion of this result is as follows: C is the minimum "radius" of a circle contains all the rows $\mathbf{r}_i$ of the channel transition matrix, where the all distances are measured with relative entropy. The center of this circle is located at $\mathbf{q}'$ which minimizes the right hand side of (8).

(Note: this is a laborious problem, and needs some care)

2. **4-ary Hamming distortion.**
A random variable X uniformly takes on values $\{0, 1, 2, 3\}$. The distortion function is the usual Hamming distortion.

$$d(x, \hat{x}) = \begin{matrix} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{matrix} \tag{9}$$

Compute the rate distortion function $R(D)$ by finding a lower bound on $I(x; \hat{x})$ and showing this lower bound to be achievable. *Hint:* Fano's inequality.

3. **Properties of optimal rate distortion code**. A good (R, D) rate distortion code with $R \approx R(D)$ puts severe constraints on the relationship of the source X^n and the representations $\hat{X}^n$. Examine the chain of inequalities (13.58–13.70) considering the conditions for equality and interpret as properties of a good code. For example, equality in (13.59) implies that $\hat{X}^n$ is a deterministic function of X^n.

4. **Ternary source with erasure distortion.**
Consider X uniformly distributed over the three inputs $\{1, 2, 3\}$, and let the distortion measure be given by the matrix

$$d(x, \hat{x}) = \begin{bmatrix} 0 & \infty & \infty & 1 \\ \infty & 0 & \infty & 1 \\ \infty & \infty & 0 & 1 \end{bmatrix} \tag{10}$$

   (a) (7 pts) Calculate the rate distortion function for this source.

   (b) (3 pts) Suggest a simple scheme to achieve any value of the rate distortion function for this source.