Homework Set no. 10

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

\[
p(y|x) = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}
\]  

(1)

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

(a) Show that for this channel

\[
\sum_i p_i D(r_i||q) \leq C
\]  

(2)

(b) Let \(q'\) be another distribution on \(Y\). Show that

\[
\sum_i p_i D(r_i||q) \leq \sum_i p_i D(r_i||q')
\]  

(3)

Use this to argue

\[
C = \max_p \sum_i p_i D(r_i||q) = \max_p \min_{q'} \sum_i p_i D(r_i||q')
\]  

(4)

and that for any distribution \(q'\),

\[
C \leq \max_i D(r_i||q').
\]  

(5)

(c) Combining the first two parts, we have

\[
\sum_i p_i D(r_i||q) \leq C \leq \min_{q'} \max_i D(r_i||q')
\]  

(6)

Now justify the following argument:

\[
\max_p \min_{q'} \sum_i p_i D(r_i||q') \geq \min_{q'} \max_i D(r_i||q')
\]  

(7)

(This is a very general result that \(\max \min \geq \min \max\), and can be argued from first principles.)
Combining the previous two equations, we have

\[ \max_{p} \sum_{i} p_i D(r_i|q) = C = \min_{q'} \max_{i} D(r_i|q') \]  \hspace{1cm} (8)

A simple geometric interpretation of this result is as follows: \( C \) is the minimum “radius” of a circle contains all the rows \( r_i \) of the channel transition matrix, where the all distances are measured with relative entropy. The center of this circle is located at \( q' \) 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.

\[ d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases} \]  \hspace{1cm} (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

\[ d(x, \hat{x}) = \begin{bmatrix} 0 & \infty & \infty & 1 \\ \infty & 0 & \infty & 1 \\ \infty & \infty & 0 & 1 \end{bmatrix} \]  \hspace{1cm} (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.