Homework no. 2

1. **Relative entropy is not symmetric:** Let the random variable \( X \) have three possible outcomes \( \{a, b, c\} \). Consider two distributions on this random variable

\[
\begin{array}{c|cc}
\text{Symbol} & p(x) & q(x) \\
\hline
a & 3/4 & 1/3 \\
b & 3/16 & 1/3 \\
c & 1/16 & 1/3 \\
\end{array}
\]

Calculate \( H(p), H(q), D(p||q) \) and \( D(q||p) \). Verify that in this case \( D(p||q) \neq D(q||p) \).

2. **Conditioning Reduces Entropy:** We showed in class that conditioning on a random variable \( Y \) reduces the entropy of a random variable \( X \):

\[
H(X | Y) \leq H(X).
\]

- Either prove or disprove the following statement:
  “conditioning on individual values of \( Y \) never increases entropy”, namely

\[
H(X | Y = y) \leq H(X) \forall y \in Y
\]

3. **Relative entropy is cost of miscoding:** In this problem, we explore a fundamental interpretation of the relative entropy.

Let the random variable \( X \) have five possible outcomes \( \{1, 2, 3, 4, 5\} \). Consider two distributions on this random variable

\[
\begin{array}{c|cc|cc}
\text{Symbol} & p(x) & q(x) & C_1(x) & C_2(x) \\
\hline
1 & 1/2 & 1/2 & 0 & 0 \\
2 & 1/4 & 1/8 & 10 & 100 \\
3 & 1/8 & 1/8 & 110 & 101 \\
4 & 1/16 & 1/8 & 1110 & 110 \\
5 & 1/16 & 1/8 & 1111 & 111 \\
\end{array}
\]

(a) Calculate \( H(p), H(q), D(p||q) \) and \( D(q||p) \).

(b) The last two columns above represent codes for the random variable. Verify that the average length of \( C_1 \) under \( p \) is equal to the entropy \( H(p) \). Thus \( C_1 \) is optimal for \( p \). Verify that \( C_2 \) is optimal for \( q \).

(c) Now assume that we use code \( C_2 \) when the distribution is \( p \). What is the average length of the codewords. By how much does it exceed the entropy \( p \)?

(d) What is the loss if we use code \( C_1 \) when the distribution is \( q \)?

4. **Proof of Theorem 3.3.1.** Let \( X_1, X_2, \ldots, X_n \) be i.i.d. \( \sim p(x) \). Let \( B_\delta^{(n)} \subset \mathcal{X}^n \) such that \( \Pr(B_\delta^{(n)}) > 1 - \delta \). Fix \( \epsilon < \frac{1}{2} \).

(a) Given any two sets \( A, B \) such that \( \Pr(A) > 1 - \epsilon_1 \) and \( \Pr(B) > 1 - \epsilon_2 \), show that \( \Pr(A \cap B) > 1 - \epsilon_1 - \epsilon_2 \). Hence \( \Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) \geq 1 - \epsilon - \delta \).
(b) Justify the steps in the chain of inequalities

\[
1 - \epsilon - \delta \leq \Pr(A^{(n)}_\epsilon \cap B^{(n)}_\delta) \leq \sum_{A^{(n)}_\epsilon \cap B^{(n)}_\delta} p(x^n) \leq \sum_{A^{(n)}_\epsilon \cap B^{(n)}_\delta} 2^{-n(H - \epsilon)} = |A^{(n)}_\epsilon \cap B^{(n)}_\delta| 2^{-n(H - \epsilon)} \leq |B^{(n)}_\delta| 2^{-n(H - \epsilon)}.
\]

(c) Complete the proof of the theorem.