## 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

| Symbol | $p(x)$ | $q(x)$ |
|---|---|---|
| a | 3/4 | 1/3 |
| b | 3/16 | 1/3 |
| c | 1/16 | 1/3 |

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 $X$:

$$\mathrm{H}(X \mid Y) \leq \mathrm{H}(X).$$

• Either prove or disprove the following statement:
"conditioning on individual values of $Y$ never increases entropy", namely

$$\mathrm{H}(X \mid Y = y) \leq \mathrm{H}(X) \forall y \in \mathcal{Y}$$

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

| Symbol | $p(x)$ | $q(x)$ | $C_1(x)$ | $C_2(x)$ |
|---|---|---|---|---|
| 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 |

(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

$$
\begin{aligned}
1-\epsilon-\delta &\le \Pr(A_\epsilon^{(n)} \cap B_\delta^{(n)}) && (1)\\
&= \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} p(x^n) && (2)\\
&\le \sum_{A_\epsilon^{(n)} \cap B_\delta^{(n)}} 2^{-n(H-\epsilon)} && (3)\\
&= |A_\epsilon^{(n)} \cap B_\delta^{(n)}| 2^{-n(H-\epsilon)} && (4)\\
&\le |B_\delta^{(n)}| 2^{-n(H-\epsilon)}. && (5)
\end{aligned}
$$

(c) Complete the proof of the theorem.