COMPSCI 650: HOMEWORK 2

INSTRUCTOR: A. MAZUMDAR

All problems carry equal points
Some problems are from Cover and Thomas, 2nd Ed.
Due Date: Mar 10, 2016: Start of Class/Midterm

(1)
- Let X be a discrete random variable. Show that the entropy of a function of X is less than or equal to the entropy of X.
- Prove the data-processing inequality for relative entropy. That is, suppose X is a (discrete) random variable and $Y = g(X)$. Let P_1 and P_2 be two distributions for X. Define the distributions $\hat{P}_1$ and $\hat{P}_2$ for Y as follows.
$$\hat{P}_1(Y = y) = P_1(g^{-1}(y)),$$
and
$$\hat{P}_2(Y = y) = P_2(g^{-1}(y)).$$
Now prove that,
$$D(P_1||P_2) \geq D(\hat{P}_1||\hat{P}_2).$$

(2) Pinsker's inequality. For any two distributions on a discrete support set $\mathcal{X}$,
$$D(P_1||P_2) \geq \frac{2}{\ln 2}\|P_1 - P_2\|_{TV}^2.$$
In the class we have shown Pinsker's inequality to be true for two Bernoulli distributions. Use that fact and the data-processing inequality to prove Pinsker's inequality for two arbitrary distributions. (Hint: For a random variable X with support $\mathcal{X}$ and $A \subset \mathcal{X}$, define a new Bernoulli random variable Y: $Y = 1$ if $X \in A$. This defines two distributions $\hat{P}_1$ and $\hat{P}_2$ for Y, where $\hat{P}_1(1) = P_1(A)$ and $\hat{P}_2(1) = P_2(A)$.)

(3) We are given the following joint distribution on (X, Y), $X \in \{1, 2, 3\}, Y \in \{1, 2, 3\}$.
$$p(x,y) = \begin{cases} \frac{1}{6}, & x = y; \\ \frac{1}{12}, & x \neq y \end{cases}$$
Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = \Pr\{\hat{X}(Y) \neq X\}$.
- Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated P_e.
- Evaluate Fanos inequality for this problem and compare.

(4) Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim p(x)$. Consider the hypothesis test $H1 : p = p_1$ vs. $H2 : p = p_2$. Let
$$p_1(x) = \begin{cases} \frac{1}{2}, & x = -1; \\ \frac{1}{4}, & x = 0; \\ \frac{1}{4}, & x = 1. \end{cases}$$

$$p_2(x) = \begin{cases} \frac{1}{4}, & x = -1; \\ \frac{1}{4}, & x = 0; \\ \frac{1}{2}, & x = 1. \end{cases}$$

Find the probability of error, $\Pr\{\text{Decide } H2|H1 \text{ true}\}$ in the best hypothesis test of H1 vs. H2 subject to $\Pr\{\text{Decide } H1|H2 \text{ true}\} \le \frac{1}{2}$.

(5) Consider an n-length random binary $\{0,1\}$-sequence $X_1^n \equiv X_1, \dots, X_n$, where each X_i is independently generated according to a Bernoulli($\frac{1}{2}$) distribution. Consider another random sequence $Y_1^n \equiv Y_1, \dots, Y_n$ exactly similarly generated. The *Hamming distance* between the two sequences, $d(X_1^n, Y_1^n)$ is defined to be the number of coordinates where the two sequences differ:

$$d(X_1^n, Y_1^n) \equiv |\{i : X_i \neq Y_i\}|.$$

- What is $\mathbb{E}d(X_i^n, Y_i^n)$, the average value of the Hamming distance between the two sequences?
- Suppose each entry of X_1^n is flipped with probability p. And as a result we obtain a sequence $\hat{X}_1^n$. What is $\mathbb{E}d(X_1^n, \hat{X}_1^n)$?
- What is $\mathbb{E}d(Y_1^n, \hat{X}_1^n)$?
- What is $Pr\{d(X_1^n, \hat{X}_1^n) \ge d(Y_1^n, \hat{X}_1^n)\}$

(6) Let $\{X_i\}$ be i.i.d. $\sim p(x), x \in \{1, 2, \dots\}$. Consider two hypotheses, $H0 : p(x) = p_0(x)$ vs. $H1 : p(x) = p_1(x)$, where $p_0(x) = (\frac{1}{2})^x$ and and $p_1(x) = qp^{x-1}, x = 1, 2, 3, \dots$

- Find $D(p_0||p_1)$.
- Let $Pr\{H0\} = \frac{1}{2}$. Find the minimal probability of error test for $H0$ vs. $H1$ given data $X_1, X_2, \dots, X_n \sim p(x)$.

(7) Let $Pr(X = i) = p_i, i = 1, 2, \dots, m$, and let $p_1 \ge p_2 \ge p_3 \ge \cdots \ge p_m$. The minimal probability of error predictor of X is $\hat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$. Maximize $H(\{p_1, \dots, p_m\})$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of H.