2nd Semester 2010

Homework Set #2
Entropy, Mutual Information, Divergence and Jensen's Inequality

1. **The value of a question.**
   Let $X \sim p(x), \quad x = 1, 2, \ldots, m$.

   We are given a set $S \subseteq \{1, 2, \ldots, m\}$. We ask whether $X \in S$ and receive the answer

   $$Y = \begin{cases} 1, & \text{if } X \in S \\ 0, & \text{if } X \notin S. \end{cases}$$

   Suppose $\Pr\{X \in S\} = \alpha$.

   (a) Find the decrease in uncertainty $H(X) - H(X|Y)$.

   (b) Is the set S with a given probability α is as good as any other.

2. **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	1/2	1/3
b	1/4	1/3
c	1/4	1/3

   Calculate $H(p), H(q), D(p \parallel q)$ and $D(q \parallel p)$.

   Verify that in this case $D(p \parallel q) \neq D(q \parallel p)$.

3. **True or False**

   Let X, Y, Z be discrete random variable. Copy each relation and write **true** or **false**. If it's true, prove it. If it is false give a counterexample or prove that the opposite is true.

   For instance:

- $H(X) \geq H(X|Y)$ is **true**. Proof: In the class we showed that $I(X;Y) > 0$, hence $H(X) - H(X|Y) > 0$.
- $H(X) + H(Y) \leq H(X,Y)$ is **false**. Actually the opposite is true, i.e., $H(X) + H(Y) \geq H(X,Y)$ since $I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0$.

(a) If $H(X|Y) = H(X)$ then X and Y are independent.

(b) For any two probability mass functions (pmf) P, Q,

$$D\left(\frac{P+Q}{2}||Q\right) \leq \frac{1}{2}D(P||Q),$$

where $D(||)$ is a divergence between two pmfs.

(c) Let X and Y be two independent random variables. Then

$$H(X+Y) \geq H(X).$$

(d) $I(X;Y) - I(X;Y|Z) \leq H(Z)$

(e) If $f(x,y)$ is a convex function in the pair (x,y), then for a fixed y, $f(x,y)$ is convex in x, and for a fixed x, $f(x,y)$ is convex in y.

(f) If for a fixed y the function $f(x,y)$ is a convex function in x , and for a fixed x, $f(x,y)$ is convex function in y, then $f(x,y)$ is convex in the pair (x,y). (Examples of such functions are $f(x,y) = f_1(x) + f_2(y)$ or $f(x,y) = f_1(x)f_2(y)$ where $f_1(x)$ and $f_2(y)$ are convex.)

4. **Random questions.**
One wishes to identify a random object $X \sim p(x)$. A question $Q \sim r(q)$ is asked at random according to $r(q)$. This results in a deterministic answer $A = A(x,q) \in \{a_1, a_2, \ldots\}$. Suppose the object X and the question Q are independent. Then $I(X;Q,A)$ is the uncertainty in X removed by the question-answer (Q,A).

(a) Show $I(X;Q,A) = H(A|Q)$. Interpret.

(b) Now suppose that two i.i.d. questions $Q_1, Q_2 \sim r(q)$ are asked, eliciting answers A_1 and A_2. Show that two questions are less valuable than twice the value of a single question in the sense that $I(X;Q_1,A_1,Q_2,A_2) \leq 2I(X;Q_1,A_1)$.

5. **Entropy bounds.**
Let $X \sim p(x)$, where x takes values in an alphabet $\mathcal{X}$ of size m. The entropy $H(X)$ is given by

$$\begin{aligned} H(X) &\equiv -\textstyle\sum_{x\in\mathcal{X}} p(x)\log p(x) \\ &= E_p \log \tfrac{1}{p(X)}. \end{aligned}$$

Use Jensen's inequality ($Ef(X) \leq f(EX)$, if f is concave) to show

(a) $H(X) \leq \log E_p \frac{1}{p(X)}$
$= \log m$.

(b) $-H(X) \leq \log(\sum_{x\in\mathcal{X}} p^2(x))$, thus establishing a lower bound on $H(X)$.

(c) Evaluate the upper and lower bounds on $H(X)$ when $p(x)$ is uniform.

(d) Let X_1, X_2 be two independent drawings of X. Find $\Pr\{X_1 = X_2\}$ and show $\Pr\{X_1 = X_2\} \geq 2^{-H}$.

6. **Bottleneck.**
Suppose a (non-stationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \to X_2 \to X_3$, $X_1 \in \{1, 2, \ldots, n\}$, $X_2 \in \{1, 2, \ldots, k\}$, $X_3 \in \{1, 2, \ldots, m\}$, and $p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$.

(a) Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.

(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.