2nd Semester 2009

Solutions to Homework Set #2
Properties of Entropy and Mutual Information

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 it true that any set S with a given probability α is as good as any other.

   Solution: The value of a question.

   (a) Consider
   $$H(X) - H(X|Y) = H(Y) - H(Y|X) = H(Y) = H_b(\alpha) \qquad (1)$$

   (b) Yes, since the answer depends only on α.

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 $D(p \parallel q) \neq D(q \parallel p)$.

   Solution: Relative entropy is not symmetric.

(a) $H(p) = 1/2 \log 2 + 2 \times 1/4 \log 4 = 1.5$ bits.

(b) $H(q) = 3 \times 1/3 \log 3 = \log 3 = 1.585$ bits.

(c) $D(p\|q) = 1/2 \log 3/2 + 2 \times 1/4 \log 3/4 = \log 3 - 3/2 = 0.0850$ bits.

(d) $D(q\|p) = 1/3 \log 2/3 + 2 \times 1/3 \log 4/3 = 5/3 - \log 3 = 0.0817$ bits.

$D(p\|q) \neq D(q\|p)$ as expected.

3. **True or False**

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

$$I(X;Y) = H(X) - H(X|Y)$$

If $I(X;Y) = 0$ then $H(X) = H(X|Y)$. We can write:

$$I(X;Y) = D(\ P_{X,Y} \ || \ P_X P_Y \ ) = 0$$

$D(Q||P) = 0$ iff $P(x) = Q(x)\ \forall x$, therefore $P_{X,Y}(x,y) = P_X(x)P_Y(y)$ for every x, y and as result $X \perp Y$.

(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.
True:
Using the concave property of the divergence function:

$$D\left(\lambda P + (1-\lambda)Q \ || \ Q\right) \leq \lambda D(P \ || \ Q) + (1-\lambda)D(Q \ || \ Q)$$

Assigning $\lambda = \frac{1}{2}$, and since $D(Q||Q) = 0$:

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

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

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

True:

$$H(X+Y) \geq H(X+Y|Y) \stackrel{(a)}{=} H(X)$$

(a) - since X is independent of Y.

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

True:

$$\begin{aligned} I(X;Y) - I(X;Y|Z) &= H(X) - H(X|Y) - [H(X|Z) - H(X|Y,Z)] \\ &= \underbrace{H(X) - H(X|Z)}_{I(X;Z)} - \underbrace{[H(X|Y) - H(X|Y,Z)]}_{\geq 0} \\ &\leq I(X;Z) \\ &= H(Z) - \underbrace{H(Z|X)}_{\geq 0} \\ &\leq H(Z) \end{aligned}$$

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 $I(X;Q_1,A_1,Q_2,A_2) \leq 2I(X;Q_1,A_1)$.

Solution: Random questions.

(a) Since A is a deterministic function of (Q, X), $H(A|Q, X) = 0$. Also since X and Q are independent, $H(Q|X) = H(Q)$. Hence,

$$\begin{aligned} I(X; Q, A) &= H(Q, A) - H(Q, A, |X) \\ &= H(Q) + H(A|Q) - H(Q|X) - H(A|Q, X) \\ &= H(Q) + H(A|Q) - H(Q) \\ &= H(A|Q). \end{aligned}$$

The interpretation is as follows. The uncertainty removed in X given (Q, A) is the same as the uncertainty in the answer given the question.

(b) Using the result from part (a) and the fact that questions are independent, we can easily obtain the desired relationship.

$$\begin{aligned} I(X; Q_1, A_1, Q_2, A_2) &\stackrel{(a)}{=} I(X; Q_1) + I(X; A_1|Q_1) + I(X; Q_2|A_1, Q_1) \\ &\quad + I(X; A_2|A_1, Q_1, Q_2) \\ &\stackrel{(b)}{=} I(X; A_1|Q_1) + H(Q_2|A_1, Q_1) - H(Q_2|X, A_1, Q_1) \\ &\quad + I(X; A_2|A_1, Q_1, Q_2) \\ &\stackrel{(c)}{=} I(X; A_1|Q_1) + I(X; A_2|A_1, Q_1, Q_2) \\ &= I(X; A_1|Q_1) + H(A_2|A_1, Q_1, Q_2) - H(A_2|X, A_1, Q_1, Q_2) \\ &\stackrel{(d)}{=} I(X; A_1|Q_1) + H(A_2|A_1, Q_1, Q_2) \\ &\stackrel{(e)}{\leq} I(X; A_1|Q_1) + H(A_2|Q_2) \\ &\stackrel{(f)}{=} 2I(X; A_1|Q_1) \end{aligned}$$

(a) Chain rule.
(b) X and Q_1 are independent.
(c) Q_2 are independent of X, Q_1, and A_1.
(d) A_2 is completely determined given Q_2 and X.
(e) Conditioning decreases entropy.
(f) Result from part (a).

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 ($Ef(X) \le f(EX)$, if f is concave) to show

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

(b) $-H(X) \le \log(\sum_{x\in\mathcal{X}} p^2(x))$, 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 $\Pr\{X_1 = X_2\} \ge 2^{-H}$.

Solution: Entropy Bounds.
To prove (a) observe that

$$\begin{aligned} H(X) &= E_p \log \frac{1}{p(X)} \\ &\le \log E_p \frac{1}{p(X)} \\ &= \log \sum_{x\in\mathcal{X}} p(x) \frac{1}{p(x)} \\ &= \log m \end{aligned}$$

where the the first inequality follows from Jensen's, and the last step follows since the size of $\mathcal{X}$ is m.

To prove (b) proceed

$$\begin{aligned} -H(X) &= E_p \log p(X) \\ &\le \log E_p p(X) \\ &= \log \left(\sum_{x\in\mathcal{X}} p^2(x) \right) \end{aligned}$$

where the second step again follows from Jensen's and the third step is just the definition of $E_p(p(X)$. Thus, we have the lower bound

$$H(X) \geq -\log\left(\sum_{x\in\mathcal{X}} p^2(x)\right) \quad .$$

The upper bound is m irrespective of the distribution. Now, $p(x) = 1/m$ for the uniform distribution, and therefore

$$\begin{aligned} -\log\sum_{x\in\mathcal{X}} p^2(x) &= -\log\sum_{x\in\mathcal{X}}\frac{1}{m^2} \\ &= -\log\frac{1}{m} \end{aligned}$$

and therefore the upper and lower bounds agree, and are $\log m$. A direct calculation of the entropy yields the same result immediately.

The derivation of (d) follows from

$$\begin{aligned} \Pr\{X_1 = X_2\} &= \sum_{x,y\in\mathcal{X}} \Pr\{X_1 = x, X_2 = y\}\delta_{xy} \\ &= \sum_{x\in\mathcal{X}} p^2(x) \end{aligned}$$

where the second step follows from the independence of X_1, X_2, and the fact that they are identically distributed $X_1, X_2 \sim p(x)$. Here δ_{xy} is Kronecker's delta function.

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.

Solution: Bottleneck.

(a) From the data processing inequality, and the fact that entropy is maximum for a uniform distribution, we get

$$
\begin{aligned}
I(X_1; X_3) &\leq I(X_1; X_2) \\
&= H(X_2) - H(X_2 \mid X_1) \\
&\leq H(X_2) \\
&\leq \log k.
\end{aligned}
$$

Thus, the dependence between X_1 and X_3 is limited by the size of the bottleneck. That is $I(X_1; X_3) \leq \log k$.

(b) For $k = 1$, $0 \leq I(X_1; X_3) \leq \log 1 = 0$ so that $I(X_1, X_3) = 0$. Thus, for $k = 1$, X_1 and X_3 are independent.