Problem Set 1

15-859 Information Theory and Applications in TCS
Name: Shashank Singh
Email: sss1@andrew.cmu.edu
Due: Thursday, February 7, 2013

Problem 1

All entropies are given in bits.

$P(Y=3) = \frac{1}{4}, P(Y=4) = P(Y=5) = \frac{3}{8}$, so $H(Y) = \frac{1}{4}\log_2(4) + 2 \cdot \frac{3}{8}\log_2\left(\frac{8}{3}\right) = \boxed{\frac{11 - 3\log_2 3}{4}}.$

Since Y is a (deterministic) function of X, $\boxed{H(Y \mid X) = 0}$ and $I(X;Y) = H(Y) = \boxed{\frac{11 - \log_2 3}{4}}.$

$H(X \mid Y = 3) = \log_2 2 = 1$, $H(X \mid Y = 4) = \log_2 6 = 1 + \log_2 3$, and $H(X \mid Y = 5) = \log_2 12 = 2 + log_2 3$ (computed by counting the number of possible series of each length). Thus,

$$H(X \mid Y) = \mathop{\mathbb{E}}_{y \in \{3,4,5\}} [H(X \mid Y = y)] = \frac{1}{4} \cdot 1 + \frac{3}{8}(3 + 2\log_2 3) = \boxed{\frac{11 + 6\log_2 3}{8}}.$$

Thus, $H(X) = H(X \mid Y) + H(Y) = \boxed{\frac{33}{8}}.$

Problem 2

(a) For $i \in \{1, 2, \ldots, n-2\}$, $I(B_i, B_{i+1} \mid B_1, B_2, \ldots, B_{i-1}) = 0$, since B_i is independent of B_{i+1} given $B_1, B_2, \ldots, B_{i-1}$. However, if $i = n-1$, then, $I(B_i, B_{i+1} \mid B_1, B_2, \ldots, B_{i-1}) = 1$, since $H(B_{i+1}) = 1$, and B_{i+1} can be uniquely determined from B_i, given the values of $B_1, B_2, \ldots, B_{i-1}$.

(b) Since conditioning cannot reduce entropy, $H(Y \mid X, Z) \leq H(Y \mid X)$.

$$\begin{aligned} H(X,Y,Z) + H(X) &= H(Y,Z \mid X) + 2H(X) \\ &= H(Y \mid X, Z) + H(Z \mid X) + 2H(X) \\ &\leq H(Y \mid X) + H(Z \mid X) + 2H(X) \\ &= H(X,Z) + H(X,Z). \end{aligned}$$

This inequality can be re-written as the "submodular" inequality. Furthermore, $H(Y \mid X, Z) = H(Y \mid X)$ if and only if Y is conditionally independent of Z given X, so that this is precisely the condition under which equality holds. ∎

Problem 3

(a) Since, for $X = x$ fixed, $Z = z$ if and only if $Y = z - x$,

$$\begin{aligned}
H(Z \mid X) &= \sum_x p(x) \sum_z p(Z = z \mid X = x) \log \left(\frac{1}{p(Z = z \mid X = x)} \right) \\
&= \sum_x p(x) \sum_z p(Y = z - x \mid X = x) \log \left(\frac{1}{p(Y = z - x \mid X = x)} \right) \\
&= \sum_x p(x) \sum_y p(Y = y \mid X = x) \log \left(\frac{1}{p(Y = y \mid X = x)} \right) = H(Y \mid X). \quad \blacksquare
\end{aligned}$$

(b) As shown in part (d) below, if $X \perp Y$, then $H(Z) = H(X) + H(Y) \geq \max\{H(X), H(Y)\}$, since entropy is non-negative. $\blacksquare$

(c) Suppose $X \sim \text{Bernoulli}\,(1/2)$ and $Y = -X$. Then, $H(X) = H(Y) = 1$, but, since Z is always 0, $H(Z) = 0 < 1 = \min\{H(X), H(Y)\}$. $\blacksquare$

(d) By part (a),

$$\begin{aligned}
H(X) + H(Y) - H(Z) &= H(X) + H(Y) - H(Z \mid X) - H(X) \\
&= H(Y) - H(Y \mid X) = I(X; Y).
\end{aligned}$$

Since $I(X; Y) = 0$ if and only if $X \perp Y$, $H(Z) = H(X) + H(Y)$ if and only if $\boxed{X \perp Y.}$ $\blacksquare$

Problem 4

(a) For $x \in \{0, 1\}^n, r \in [0, \infty)$, let $B(x, r) \subseteq \{0, 1\}^n$ denote the ball of Hamming radius r centered at x. By the given inequality, the cardinality of $B(x, \tau n)$ is

$$|B(x, \tau n)| = \sum_{j=0}^{\tau n} \binom{n}{j} \leq 2^{h(\tau) n} \tag{1}$$

(there are $\binom{n}{j}$ strings with Hamming distance exactly j from x, since we construct such a string by choosing j bits of x to flip). If C is a τ-covering, then $\{0, 1\}^n = \bigcup_{x \in C} B(x, \tau n)$. Thus,

$$\begin{aligned}
2^n = |\{0, 1\}^n| = \left| \bigcup_{x \in C} B(x, \tau n) \right| &\leq \sum_{x \in C} |B(x, \tau n)| && \text{(by the Union Bound)} \\
&\leq \sum_{x \in C} 2^{h(\tau) n} = |C| 2^{h(\tau) n}, && \text{(by (1))}
\end{aligned}$$

which can be rewritten in the desired form:

$$|C| \geq \frac{2^n}{2^{h(\tau) n}} = 2^{(1 - h(\tau)) n}. \quad \blacksquare$$

(b) Couldn't get this one. ☹

Problem 5

(a) Clearly the leaves of the entire tree are $\{a_1, a_2, \ldots, a_n\}$. Furthermore, if the leaves in the subtree rooted at some internal node N are $\{a_i, a_{i+1}, \ldots, a_j\}$ $(1 \leq i < j \leq n)$, then, since for some $k \in \{i, i+1, \ldots, j\}$, the leaves in the subtrees rooted at the left and right children of N are $\{i, i+1, \ldots, k\}$ and $\{k+1, k+2, \ldots, j\}$, so that both children of N have the desired property. Thus, by induction, the desired property holds for all internal nodes in the tree. ■

(b) In the sum $\sum_{[i,j]\in\mathcal{I}} q_{[i,j]}$, each p_i appears once for each internal node which is an ancestor of $[i,i]$. Since the length l_i of the code for a_i is the number of ancestors of $[i,i]$,

$$\sum_{[i,j]\in\mathcal{I}} q_{[i,j]} = \sum_{i\in\{1,2,\ldots,n\}} p_i l_i = L. \quad ■$$

(c) For any node $[i,j]$, let $S([i,j])$ denote the set of nodes in the subtree rooted at $[i,j]$. We show by induction that, for each internal node $[i,j]$ in the tree,

$$H(X \mid X \in \{a_i, a_{i+1}, \ldots, a_j\}) = \sum_{[i',j']\in S([i,j])} \frac{q_{[i',j']}}{q_{[i,j]}} h\left(\frac{q_{[i',k']}}{q_{[i',j']}}\right),$$

which, in the case $i = 1, j = n$, reduces to the desired result. If $i = j$, this is trivial, since $H(X \mid X = a_i) = 0 = h(1)$. Suppose now, that the result holds for the left and right children of some internal node $[i,j]$. Then, letting D be a Bernoulli random variable with $D = L$ if $X \in \{a_i, a_{i+1}, \ldots, a_k\}$ and $D = R$ otherwise, computing conditional entropy as an expected value

$$\begin{aligned}
H(X \mid X \in \{a_i, \ldots, a_j\}) &= H(X \mid L, X \in \{a_i, \ldots, a_j\}) + H(L) \\
&= \frac{q_{[i,k]}}{q_{[i,j]}} \left(\sum_{[i',j']\in S([i,k])} \frac{q_{[i',j']}}{q_{[i,k]}} h\left(\frac{q_{[i',k']}}{q_{[i',j']}}\right) \right) \\
&+ \frac{q_{[k+1,j]}}{q_{[i,j]}} \left(\sum_{[i',j']\in S([k+1,j])} \frac{q_{[i',j']}}{q_{[k+1,j]}} h\left(\frac{q_{[i',k']}}{q_{[i',j']}}\right) \right) + \frac{q_{[i,j]}}{q_{[i,j]}} h\left(\frac{q_{[i,k]}}{q_{[i,j]}}\right) \\
&= \sum_{[i',j']\in S([i,j])} \frac{q_{[i',j']}}{q_{[i,j]}} h\left(\frac{q_{[i',k']}}{q_{[i',j']}}\right),
\end{aligned}$$

so that the result holds for $[i,j]$. ■

(d) By the results of parts (b) and (c),

$$\begin{aligned}
L - H(X) &= \left(\sum_{[i,j]\in\mathcal{I}} q_{[i,j]}\right) - \sum_{[i,j]\in\mathcal{I}} q_{[i,j]} h\left(\frac{q_{[i,k]}}{q_{[i,j]}}\right) \\
&\leq \sum_{[i,j]\in\mathcal{I}} q_{[i,j]} - 2\left(\frac{\min\{q_{[i,k]}, q_{[k+1,j]}\}}{q_{[i,j]}}\right) \quad \text{(since } h(1-x) = h(x) \text{ and } h(x) \geq 2x\text{)} \\
&= \sum_{[i,j]\in\mathcal{I}} \left|q_{[k+1,j]} - q_{[i,k]}\right|. \quad \blacksquare \qquad \text{(since } q_{[i,j]} = q_{[i,k]} + q_{[k+1,j]}\text{)}
\end{aligned}$$

(e) Suppose, for sake of contradiction, that, for some $[i,j] \in \mathcal{I}$, $\left|q_{[i,k]} - q_{[k+1,j]}\right| > \max\{p_k, p_{k+1}\}$. If $q_{[i,k]} < q_{[k+1,j]}$,

$$\left|q_{[i,k+1]} - q_{[k+2,j]}\right| = \left|q_{[i,k]} + p_{k+1} - (q_{[k+1,j]} - p_{k+1})\right| < \left|q_{[i,k]} - q_{[k+1,j]}\right|,$$

and, if $q_{[i,k]} > q_{[k+1,j]}$

$$\left|q_{[i,k-1]} - q_{[k,j]}\right| = \left|q_{[i,k]} - p_k - (q_{[k+1,j]} + p_k)\right| < \left|q_{[i,k]} - q_{[k+1,j]}\right|,$$

Either case contradicts to choice of $k = \operatorname{argmax}_{\ell: i \leq \ell < j} \left|q_{[i,k]} - q_{[k+1,j]}\right|$. $\blacksquare$

(f) By parts (d) and (e), since each p_i can be used at most twice (once as p_k and once as p_{k+1})

$$L - H(X) \leq \sum_{[i,j]\in\mathcal{I}} \left|q_{[i,k]} - q_{[k+1,j]}\right| \leq \sum_{[i,j]\in\mathcal{I}} \max\{p_k, p_{k+1}\} \leq 2. \quad \blacksquare$$

Problem 6

Pinsker's Inequality can be rewritten in the form

$$\sqrt{2D(p \,||\, q)} \geq \sum_{a\in A} |p(a) - q(a)|. \tag{2}$$

Thus, since mutual information is the divergence of joint and product distributions (shown in class),

$$\begin{aligned}
\sqrt{2I(X;Y)} &= \sqrt{2D(p(x,y); p(x)p(y))} \\
&\geq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} |p(x,y) - p(x)p(y)| && \text{(by (2))} \\
&= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} |p(y\,|\,x)p(x) - p(x)p(y)| && \text{(definition of conditional probability)} \\
&= \sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} |p(y\,|\,x) - p(y)| \\
&= \sum_{x\in\mathcal{X}} p(x)d(x) = \mathop{\mathbb{E}}_{x\leftarrow\mathcal{X}} [d(x)]. \quad \blacksquare && \text{(definitions of } d\text{, expected value)}
\end{aligned}$$