EE6340 - Information Theory

Problem Set 2 Solution

February 21, 2013

1. a) From Z=X+Y, $\mathbb{P}(Z = z|X = x) = \mathbb{P}(Y = Z - x|X = x)$

$$\begin{aligned}
H(Z|X) =& \sum_x p(x) H(Z|X = x) \\
=& - \sum_x p(x) \sum_z p(Z = z|X = x) \log_2 p(Z = z|X = x) \\
=& - \sum_x p(x) \sum_z p(Y = z - x|X = x) \log_2 p(Y = z - x|X = x) \\
=& \sum_x p(x) H(Y|X = x) \\
=& H(Y|X)
\end{aligned}$$

If X and Y are independent, $H(Y|X) = H(Y)$.
Also, $H(Z|X) \leq H(Z)$ (conditioning reduces entropy).
$\therefore H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$
$\implies H(Z) \geq H(Y)$. Similarly, we can prove that $H(Z) \geq H(X)$.

b) Consider the two random variables X and Y such that $\mathbb{P}(X = 0) = 0.5$, $\mathbb{P}(X = 1) = 0.5$ and $X = -Y$(dependent). So, $H(X) = H(Y) = 1$ bit, while $H(Z) = 0$ since $\mathbb{P}(z = 0) = 1$.

c) $H(Z) \leq H(X, Y) \leq H(X) + H(Y)$.
This is because Z is a function of X and Y and $I(X; Y) \geq 0$. Both equalities are satisfied if Z is a bijection from (X,Y)($\implies H(Z) = H(X, Y)$) and X and Y are independent($\implies H(X, Y) = H(X) + H(Y)$).

2. a) We use algebra of entropies for the proof. Since X_1 and X_2 have disjoint support sets, define a function of X,
$\theta = f(x) = \begin{cases} 1 & when X = X_1 \\ 2 & when X = X_2 \end{cases}$

$$\begin{aligned}
H(X) =& H(X, f(X)) = H(\theta) + H(X|\theta) \\
=& H(\theta) + p(\theta = 1)H(X|\theta = 1) + p(\theta = 2)H(X|\theta = 2) \\
=& H(\alpha) + \alpha H(X_1) + (1 - \alpha)H(X_2)
\end{aligned}$$

where $H(\alpha) = -\alpha \log_2 \alpha - (1 - \alpha) \log_2(1 - \alpha)$

b) To maximise over α,
$\frac{dH(X)}{d\alpha} = 0$

$\implies$ we get $\alpha_{max} = \frac{2^{H(X_1)}}{2^{H(X_1)}+2^{H(X_2)}}$
Substituting α_{max} in $H(X)$, we get
$H_{max}(X) = \log(2^{H(X_1)} + 2^{H(X_2)})$
$\implies H(X) \leq H_{max}(X) = \log(2^{H(X_1)} + 2^{H(X_2)})$
$\therefore 2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$
Thus the effective alphabet sizes add if α is chosen as α_{max}.

c) Since X_1 and X_2 are Uniformly distributed, $H(X_1) = \log m$ and $H(X_2) = \log(n-m)$.
$\therefore \alpha_{max} = \frac{m}{n}$ and $H_{max}(X) = \log(2^{H(X_1)} + 2^{H(X_2)}) = \log(m+n-m) = \log n$

3. Let $P_1 = \{p_1, p_2, ..., p_i, ..., p_j, ...p_m\}$ and
$P_2 = \{p_1, p_2, ..., \frac{p_i+p_j}{2}, ..., \frac{p_i+p_j}{2}, ...p_m\}$

$$\begin{aligned} H(P_2) - H(P_1) &= -2(\frac{p_i+p_j}{2})\log_2(\frac{p_i+p_j}{2}) + p_i \log_2 p_i + p_j \log_2 p_j \\ &= -(p_i+p_j)\log_2(\frac{p_i+p_j}{2}) + p_i \log_2 p_i + p_j \log_2 p_j \end{aligned}$$

Log-sum inequality $\implies \sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq (\sum a_i) \log \frac{\sum a_i}{\sum b_i}$

$\implies H(P_2) - H(P_1) \geq -p_i \log_2 p_i - p_j \log_2 p_j + p_i \log_2 p_i + p_j \log_2 p_j = 0$

$\therefore H(P_2) \geq H(P_1)$.

Any transfer of probability that makes the distribution more uniform increases the entropy.

4. Since the run-lengths are functions of $X_1, X_2, ...X_n$, we can say $H(R) \leq H(X)$.
Any one X_i together with the run-lengths determines the entire sequence $X_1, X_2, ...X_n$.
Hence,

$$\begin{aligned} H(X_1, X_2, ...X_n) &= H(X_i, R) \\ &= H(R) + H(X_i|R) \\ &\leq H(R) + H(X_i) \\ &\leq H(R) + 1 \end{aligned}$$

5. a) Example for $I(X;Y|Z) < I(X;Y)$:
X is a binary Random variable and Y=X, Z=Y. In this case,
$I(X;Y) = H(X) - H(X|Y) = 1 - 0 = 1$ and
$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = 0 - 0 = 0 \implies \geq I(X;Y|Z) < I(X;Y)$

b) Example for $I(X;Y|Z) > I(X;Y)$:
As in Problem 1, consider two binary independent random variables X,Y such that Z=X+Y.
$\implies I(X;Y) = 0$
But, $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(X|Z) - 0 = H(X|Z) = \frac{1}{2} \implies I(X;Y|Z) > I(X;Y)$

6. By chain rule,
$I(X_1; X_2, X_3, ...X_n) = I(X_1; X_2) + I(X_1; X_3|X_2) + ... + I(X_1; X_n|X_2, X_3, ...X_{n-1})$
By the property of Markov chain, given the present,past and future are independent. So, all terms in the above equation except the first one are 0.

$\implies I(X_1; X_2, X_3, ...X_n) = I(X_1; X_2)$