Solutions to Homework # 2

ECE 6605 Information Theory — 9/15/03
Prof. Steven W McLaughlin

1. a) From the data processing theorem we know $I(X;Z) \leq I(X;Y)$ then

$$\begin{aligned} I(X;Z) &\leq I(X;Y) = H(Y) - H(Y|X) \\ &\leq H(Y) \text{ since } H(Y|X) \geq 0 \\ &\leq \log_2 k \end{aligned}$$

   b) If $k = 1$ then $\log_2 1 = 0$; therefore $I(X;Z) \leq 0$, but since $I(X;Z) \geq 0$ then $I(X;Z) = 0$. Which confirms that no dependence can survive the bottleneck

2. We are given $\mathbf{X}$ and can form $\mathbf{R}$ from it, therefore $\mathbf{R}$ is a function of $\mathbf{X}$. Then from problem 4 in HW. 1 we get $H(\mathbf{R}) \leq H(\mathbf{X})$.

   If we are given $\mathbf{R}$ and X_n (or any X_i for that matter), then we can fully specify $\mathbf{X}$, this tells us that $H(\mathbf{X}) = H(X_n, \mathbf{R})$. Expanding this we get

$$\begin{aligned} H(\mathbf{X}) &= H(X_n, \mathbf{R}) \\ &= H(\mathbf{R}) + H(X_n|\mathbf{R}) \\ &\leq H(\mathbf{R}) + H(X_n) \text{ since conditioning reduces entropy} \\ &\leq H(\mathbf{R}) + 1 \end{aligned}$$

   From the two results above we notice $H(\mathbf{R}) \leq H(\mathbf{X}) \leq H(\mathbf{R}) + 1$.

3. First notice that for $k \leq n-1$ because of the independence of the X_i's we get that $I(X_{k-1};X_k|X_1,X_2,\ldots,X_{k-2}) = 0$, we can gain no information about X_{k-1} from knowing $X_k|X_1,X_2,\ldots,X_{k-2}$ since we do not know the rest of the sequence.

   Now for $I(X_{n-1};X_n|X_1,\ldots,X_{n-2})$ we have the following

$$\begin{aligned} I(X_{n-1};X_n|X_1,\ldots,X_{n-2}) &= H(X_n|X_1,\ldots,X_{n-2}) - H(X_n|X_1,\ldots,X_{n-1}) \\ &= 1 - 0 \\ &= 1 \text{ bit} \end{aligned}$$

   where we used the fact that the uncertainty of X_n when we know all other X_i's is zero, since it is completely specified and the uncertainty of X_n when we know all but X_{n-1} is equal to one bit.

4. For this one follow a process similar to the proof of the AEP.

$$\begin{aligned} \lim_{n\to\infty}[p(X_1,X_2,\ldots,X_n)]^{1/n} &= \lim_{n\to\infty} 2^{\log_2 [p(X_1,X_2,\ldots,X_n)]^{1/n}} \\ &= 2^{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \log_2 p(X_i)} \\ &= 2^{E(\log_2 p(X))} \\ &= 2^{-H(X)} \end{aligned}$$

   where the second to last equality follows from the WLLN

5. a) For this one start by evaluating the limit stated

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log_2 q(X_1,X_2,\ldots,X_n) &= \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\log_2 q(X_i) \\
&= E_p(-\log_2 q(X)) \text{ convergence in probability} \\
&= \sum p(X)\log_2\frac{1}{q(X)} \text{ adding } 0 = \log_2\tfrac{p(X)}{p(X)} \\
&= \sum p(X)\log_2\frac{p(X)}{q(X)} - \sum p(X)\log_2 p(X) \\
&= D(p||q) + H(X)
\end{aligned}$$

b) Notice that there is no negative sign here, this will be important in the final answer.

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\log_2\frac{q(X_1,X_2,\ldots,X_n)}{p(X_1,X_2,\ldots,X_n)} &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{q(X_i)}{p(X_i)} \\
&= E_p(\log_2\frac{q(X)}{p(X)}) \text{ conv. in prob} \\
&= \sum p(x)\log_2\frac{q(X)}{p(X)} \\
&= -\sum p(X)\log_2\frac{p(X)}{q(X)} \\
&= -D(p||q)
\end{aligned}$$

relating this last result with the limit we get that $q(X^n) \approx p(X^n)2^{-nD(p||q)}$ for almost every sequence $X^n = (X_1,X_2,\ldots,X_n)$.