ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 10 Information Theory and Coding
Solutions to homework 5 Oct. 22, 2013

PROBLEM 1.

(a) Since the $X_1, \dots, X_n$ are i.i.d., so are $q(X_1), q(X_2), \dots, q(X_n)$, and hence we can apply the strong law of large numbers to obtain

$$\begin{aligned}\lim -\frac{1}{n}\log q(X_1,\dots,X_n) &= \lim -\frac{1}{n}\sum \log q(X_i)\\ &= -E[\log q(X)] \quad \text{w.p. } 1\\ &= -\sum p(x)\log q(x)\\ &= \sum p(x)\log\frac{p(x)}{q(x)} - \sum p(x)\log p(x)\\ &= D(p||q) + H(X).\end{aligned}$$

(b) Again, by the strong law of large numbers,

$$\begin{aligned}\lim -\frac{1}{n}\log\frac{q(X_1,\dots,X_n)}{p(X_1,\dots,X_n)} &= \lim -\frac{1}{n}\sum\log\frac{q(X_i)}{p(X_i)}\\ &= -E\Big[\log\frac{q(X)}{p(X)}\Big] \quad \text{w.p. } 1\\ &= -\sum p(x)\log\frac{q(x)}{p(x)}\\ &= \sum p(x)\log\frac{p(x)}{q(x)}\\ &= D(p||q).\end{aligned}$$

PROBLEM 2.

(a) It is easy to check that W is an i.i.d. process but Z is not. As W is i.i.d. it is also stationary. We want to show that Z is also stationary. To show this, it is sufficient to prove that the distribution of the process does not change by shift in the time domain.

$$\begin{aligned}&p_Z(Z_m = a_m, Z_{m+1} = a_{m+1}, \cdots, Z_{m+r} = a_{m+r})\\ &= \frac{1}{2}p_X(X_m = a_m, X_{m+1} = a_{m+1}, \cdots, X_{m+r} = a_{m+r})\\ &+ \frac{1}{2}p_Y(Y_m = a_m, Y_{m+1} = a_{m+1}, \cdots, Y_{m+r} = a_{m+r})\\ &= \frac{1}{2}p_X(X_{m+s} = a_m, X_{m+s+1} = a_{m+1}, \cdots, X_{m+s+r} = a_{m+r})\\ &+ \frac{1}{2}p_Y(Y_{m+s} = a_m, Y_{m+s+1} = a_{m+1}, \cdots, Y_{m+s+r} = a_{m+r})\\ &= p_Z(Z_{m+s} = a_m, Z_{m+s+1} = a_{m+1}, \cdots, Z_{m+s+r} = a_{m+r}),\end{aligned}$$

where we used the stationarity of the X and Y processes. This shows the invariance of the distribution with respect to the arbitrary shift s in time which implies stationarity.

(b) For the Z process we have

$$\begin{aligned} H(Z) &= \lim_{n\to\infty} \frac{1}{n} H(Z_1, \cdots, Z_n) \\ &= \lim_{n\to\infty} H(Z_1, \cdots, Z_n \mid \Theta) \\ &= \frac{1}{2} H(X_0) + \frac{1}{2} H(Y_0) = 1. \end{aligned}$$

W process is an i.i.d process with the distribution $p_W(a) = \frac{1}{2}p_X(a) + \frac{1}{2}p_Y(a)$. From concavity of the entropy, it is easy to see that $H(W) = H(W_0) \geq \frac{1}{2}H(X_0) + \frac{1}{2}H(Y_0) = 1$. Hence, the entropy rate of W is greater than the entropy rate of Z and the equality holds if and only if X_0 and Y_0 have the same probability distribution function.

Problem 3. (a) We can write the following chain of inequalities:

$$\begin{aligned} Q^n(\mathbf{x}) &\overset{1}{=} \prod_{i=1}^{n} Q(x_i) \overset{2}{=} \prod_{a\in\mathcal{X}} Q(a)^{N(a|\mathbf{x})} \overset{3}{=} \prod_{a\in\mathcal{X}} Q(a)^{nP_{\mathbf{x}}(a)} = \prod_{a\in\mathcal{X}} 2^{nP_{\mathbf{x}}(a)\log Q(a)} \\ &= \prod_{a\in\mathcal{X}} 2^{n(P_{\mathbf{x}}(a)\log Q(a) - P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a) + P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a))} \\ &= 2^{n\sum_{a\in\mathcal{X}}(-P_{\mathbf{x}}(a)\log\frac{P_{\mathbf{x}}(a)}{Q(a)} + P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a))} = 2^{n(-D(P_{\mathbf{x}}||Q + H(P_{\mathbf{x}}))}, \end{aligned}$$

where 1 follows because the sequence is i.i.d., grouping symbols gives 2, and 3 is the definition of type.

(b) Upper bound: We know that

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

Consider one term and set $p = k/n$. Then,

$$1 \geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} = \binom{n}{k} 2^{n\left(\frac{k}{n}\log\frac{k}{n} + \frac{n-k}{n}\log\frac{n-k}{n}\right)} = \binom{n}{k} 2^{-nh_2(\frac{k}{n})}$$

Lower bound: Define $S_j = \binom{n}{j} p^j (1-p)^{n-j}$. We can compute

$$\frac{S_{j+1}}{S_j} = \frac{n-j}{j+1}\frac{p}{1-p}.$$

One can see that this ratio is a decreasing function in j. It equals 1, if $j = np + p - 1$, so $\frac{S_{j+1}}{S_j} < 1$ for $j = \lfloor np + p \rfloor$ and $\frac{S_{j+1}}{S_j} \geq 1$ for any smaller j. Hence, S_j takes its maximum value at $j = \lfloor np + p \rfloor$, which equals k in our case. From this we have that

$$\begin{aligned} 1 = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} &\leq (n+1) \max_j \binom{n}{j} p^j (1-p)^j \\ &\leq (n+1) \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} = (n+1)\binom{n}{k} 2^{-nh_2(\frac{k}{n})}. \end{aligned}$$

The last equality comes from the derivation we had when proving the upper bound.

PROBLEM 4. Upon noticing $0.9^6 > 0.1$, we obtain $\{1, 01, 001, 0001, 00001, 000001, 0000001, 0000000\}$ as the dictionary entries.

PROBLEM 5. Since the words of a valid and prefix condition dictionary reside in the leaves of a full tree, the Kraft inequality must be satisfied with equality: Consider climbing up the tree starting from the root, choosing one of the D branches that climb up from a node with equal probability. The probability of reaching a leaf at depth l_i is then D^{-l_i}. Since the climbing process will certainly end in a leaf, we have

$$1 = \Pr(\text{ending in a leaf}) = \sum_i D^{-l_i}.$$

If the dictionary is valid but not prefix-free, by removing all words that already have a prefix in the dictionary we would obtain a valid prefix-free dictionary. Since this reduced dictionary would satisfy the Kraft inequality with equality, the extra words would cause the inequality to be violated.

PROBLEM 6.

(a) Let I be the set of intermediate nodes (including the root), let N be the set of nodes except the root and let L be the set of all leaves. For each $n \in L$ define $A(n) = \{m \in N : m \text{ is an ancestor of } n\}$ and for each $m \in N$ define $D(m) = \{n \in L : n \text{ is a descendant of } m\}$. We assume each leaf is an ancestor and a descendant of itself. Then

$$\begin{aligned} E[\text{distance to a leaf}] &= \sum_{n\in L} P(n) \sum_{m\in A(n)} d(m) \\ &= \sum_{m\in N} d(m) \sum_{n\in D(m)} P(n) = \sum_{m\in N} P(m) d(m). \end{aligned}$$

(b) Let $d(n) = -\log Q(n)$. We see that $-\log P(n_j)$ is the distance associated with a leaf. From part (a),

$$\begin{aligned} H(\text{leaves}) &= E[\text{distance to a leaf}] \\ &= \sum_{n\in N} P(n) d(n) \\ &= -\sum_{n\in N} P(n) \log Q(n) \\ &= -\sum_{n\in N} P(\text{parent of } n) Q(n) \log Q(n) \\ &= -\sum_{m\in I} P(m) \sum_{n:\ n \text{ is a child of } m} Q(n) \log Q(n) \\ &= \sum_{m\in I} P(m) H_{m'} \end{aligned}$$

(c) Since all the intermediate nodes of a valid and prefix condition dictionary have the same number of children with the same set of Q_n, each $H_n = H$. Thus $H(\text{leaves}) = H \sum_{n\in I} P(n) = HE[L]$.