ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 11 Information Theory and Coding

Solutions to homework 5 Oct 23, 2012

Problem 1. (a) We can write the following chain of inequalities:

$$Q^n(\mathbf{x}) \stackrel{1}{=} \prod_{i=1}^{n} Q(x_i) \stackrel{2}{=} \prod_{a\in\mathcal{X}} Q(a)^{N(a|\mathbf{x})} \stackrel{3}{=} \prod_{a\in\mathcal{X}} Q(a)^{nP_{\mathbf{x}}(a)} = \prod_{a\in\mathcal{X}} 2^{nP_{\mathbf{x}}(a)\log Q(a)} \tag{1}$$

$$= \prod_{a\in\mathcal{X}} 2^{n(P_{\mathbf{x}}(a)\log Q(a) - P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a) + P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a))} \tag{2}$$

$$= 2^{n\sum_{a\in\mathcal{X}}(-P_{\mathbf{x}}(a)\log\frac{P_{\mathbf{x}}(a)}{Q(a)} + P_{\mathbf{x}}(a)\log P_{\mathbf{x}}(a))} = 2^{n(-D(P_{\mathbf{x}}||Q)+H(P_{\mathbf{x}}))},$$

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

$$1 = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} \leq (n+1)\max_j \binom{n}{j} p^j (1-p)^j$$

$$\leq (n+1)\binom{n}{k}\left(\frac{k}{n}\right)^k \left(1-\frac{k}{n}\right)^{n-k} = \binom{n}{k} 2^{-nh_2(\frac{k}{n})}. \tag{3}$$

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

Problem 2. Upon noticing $0.9^6 > 0.1$, we obtain $\{1, 01, 001, 0001, 00001, 000001, 0000001, 0000000\}$ as the dictionary entries.

Problem 3. 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 4.

(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 $H(\text{leaves}) = H \sum_{n \in I} P(n) = HE[L]$.

Problem 5. (a)

$$E[F_n] = E[F_0 X_0 X_1 \ldots X_n] = F_0 (E[X_1])^n = F_0 (9/8)^n$$

We exploited the i.i.d. property of the sequence. One can see that $E[F_n] \to \infty$ with $n \to \infty$.

(b)

$$l_n = E[\log_2 F_n] = E[\log_2(F_0 X_0 X_1 \dots X_n)] = E\left[\log_2 F_0 + \sum_{i=1}^{n} \log_2 X_i\right] =$$
$$= E[\log_2 F_0] + nE[\log_2 X_1] = \log_2 F_0 - \frac{n}{2}. \tag{4}$$

(c) It concentrates around 2^{l_n}. F_n in itself is not a sum of i.i.d. variables. Taking its logarithm results such a sum, so the law of large numbers applies.

$$\log_2 F_n = \log_2 F_0 + \sum_{i=1}^{n} \log_2 X_i \to \log_2 F_0 + nE[\log X_1] = \log_2 F_0 - \frac{n}{2}.$$

(d) From the previous result it follows that although it seems appealing that the expected value of our fortune goes to infinity, it actually converges to 0 (very rapidly).

(e) We can equivalently say that instead of playing n times, we create 2^n portions of our initial money, $\binom{n}{i}$ portions of size $F_0 r^{n-i}(1-r)^i$, for all $i = 0, \dots, n$. Then we bet i times every $F_0 r^{n-i}(1-r)^i$ portion. We have seen that the more we play, the more we lose, so we should give smaller portions to large i values. The best is to set $i = 0$ for all our money, that is $r = 1$, i.e. we don't play at all.