ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 10 Introduction to Communication Systems
Solution 5 October 26, 2010, SG1 – 15:15-17:00

Problem 1. (a) Let $y_n = p(x_1, \cdots, x_n)^{\frac{1}{n}}$. Since $X_1, X_2, \cdots$ is an i.i.d. sequence, we have $p(x_1, \cdots, x_n) = \prod_{i=1}^{n} p(x_i)$ and

$$\begin{aligned}
\log y_n &= \frac{1}{n} p(x_1, \cdots, x_n) \\
&= \frac{1}{n} \log \prod_{i=1}^{n} p(x_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \log p(x_i) \\
&\overset{\text{in prob.}}{\to} E(\log p(x)) = -H(X),
\end{aligned}$$

where the last statement is due to the fact that the average of n i.i.d. samples of a random variable converges in probability to the expectation of the random variable. As a result, since $\log y_n$ converges in probability to $-H(X)$, y_n itself converges in probability to $2^{-H(X)}$.

(b) If we go along the same lines as part (a), assuming $y_n = (\prod_{i=1}^{n} f(x_i))^{\frac{1}{n}}$ we obtain

$$\log y_n = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i) \to E(\log f(x)).$$

Thus $y_n \to 2^{E(log f(X))}$. The solustion for the second part is exactly similar with the same final answer(note that $\lim_{n\to\infty} \frac{n}{n+\log n} = 1$).

(c) Let $g(u) = u^{\frac{1}{n}}$. Firstly we have $g''(u) = \frac{1}{n}(\frac{1}{n} - 1)(u)^{\frac{1}{n}-2} \leq 0$. As a result g is a concave function. Thus given a random variable Y, by Jensen's inequality we have

$$E(g(Y)) \leq g(E(Y)).$$

Now if we take $Y = \prod_{i=1}^{n} f(x_i)$, we have

$$\begin{aligned}
E(g(Y)) &= E((\prod_{i=1}^{n} f(x_i))^{\frac{1}{n}}) \\
&\leq g(E(Y)) \\
&= (E(\prod_{i=1}^{n} f(x_i)))^{\frac{1}{n}} \\
&= (\prod_{i=1}^{n} E(f(x_i)))^{\frac{1}{n}} \\
&= E(f(X)).
\end{aligned}$$

Note that this inequality holds for any $n \in \mathbb{N}$ and we have not considered the convergence issues.

Problem 2. (a) It is straightforward.

(b) It's $H(X_2|X_1)$.

(c) Note that the process is a (first-order) Markov chain since the the probability of being in each state (building) for the next time only depends on the current state (building). The transition matrix for this process would be

$$P = \begin{array}{c} \\ \text{IN} \\ \text{CO} \\ \text{SG} \end{array} \begin{array}{c} \begin{array}{ccc} \text{IN} & \text{CO} & \text{SG} \end{array} \\ \begin{pmatrix} 0 & 2/3 & 1/3 \\ 2/6 & 2/6 & 2/6 \\ 1/3 & 2/3 & 0 \end{pmatrix} \end{array},$$

where P_{ij} is the probability of going to state j given that we are in state i.

(d) The stationary distribution is a vector $\Pi = (\Pi_{\text{IN}} \quad \Pi_{\text{CO}} \quad \Pi_{\text{SG}}) = (p_1, p_2, p_3)$, where $\Pi P = \Pi$.

$$\begin{aligned}
&\frac{1}{3}p_2 + \frac{1}{3}p_3 = p_1 \\
&\frac{2}{3}p_1 + \frac{1}{3}p_2 + \frac{2}{3}p_3 = p_2 \\
&\frac{1}{3}p_1 + \frac{1}{3}p_2 = p_3 \\
&p_1 + p_2 + p_3 = 1
\end{aligned}$$

$$\begin{aligned}
&p_2 + p_3 = 3p_1 \\
&2p_1 + p_2 + 2p_3 = 3p_2 \\
&p_1 + p_2 = 3p_3 \\
&p_1 + p_2 + p_3 = 1
\end{aligned}$$

$\Rightarrow \Pi = \left(\frac{1}{4} \quad \frac{1}{2} \quad \frac{1}{4}\right)$.

(e)

$$\begin{aligned}
\mathcal{H}(X) &= \lim_{n\to\infty} \frac{1}{n} H(X_1 ... X_n) \\
&\overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_1 ... X_{i-1}) \\
&\overset{(b)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_{i-1}) \\
&\overset{(c)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_2 | X_1) \\
&= \lim_{n\to\infty} H(X_2 | X_1) \\
&= H(X_2 | X_1)
\end{aligned}$$

where in (a) the joint entropy is expanded using the chain rule, (b) is by using the property of the Markov chain, and in (c) the stationarity of the process has been used.

$$\mathcal{H}(X) = H(X_2|X_1) = \sum_{x\in\{IN,CO,SG\}} p(x)\, H(X_2|X_1 = x)$$

where $p(x)$ is the stationary distribution of the process.

$$\begin{aligned}
H(X_2|X_1 = IN) &= -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = -\frac{2}{3} + \log 3\\
H(X_2|X_1 = CO) &= -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} = \log 3\\
H(X_2|X_1 = SG) &= -\frac{2}{3} + \log 3 \quad (similar \;\; to \;\; IN \;\; case)
\end{aligned}$$

$$\mathcal{H}(X) = 2\frac{1}{4}\left(-\frac{2}{3} + \log 3\right) + \frac{1}{2}\log 3 = -\frac{1}{3} + \log 3 \cong 1.25$$

(f) The entropy of the process is the entropy of its stationary distribution.

$$H(X) = 2 - \frac{1}{4}\log\frac{1}{1/4} + \frac{1}{2}\log\frac{1}{1/2} = 1.5$$

The same relationship always holds since

$$H(X) = H(X_2) \geq H(X_2|X_1) = \mathcal{H}(X),$$

and the inequality holds because conditioning reduces the *average* entropy.
Remark: Note that it is possible that conditioning on a specific realization of a random variable *decreases* the entropy, i.e

$$H(X) < H(X|Y = y).$$

However, conditioning always reduces the *average* entropy, i.e.

$$H(X) \geq H(X|Y) = \sum_{y\in Y} p(y)\, H(X|Y = y).$$

Problem 3. (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

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

(b) Consider any leaf node say n_j. Consider the unique path in the tree from the leaf node $n_j$to the root. Let us label the nodes, which we encounter along the path to the root, as $n_j^1, n_j^2, \cdots, n_j^l$where n_j^l is the root of the tree. We observe that

$$P(n_j) = \frac{P(n_j)}{P(n_j^1)} \frac{P(n_j^1)}{P(n_j^2)} \cdots \frac{P(n_j^{l-1})}{P(n_j^l)} \tag{1}$$

where $P(n_j^i)$ are the probabilities assigned in the usual way to the intermediate nodes. Also note that $P(n_j^l) = P(\text{root}) = 1$. Thus from the definition of $Q(n)$ we can say that

$$P(n_j) = Q(n_j)Q(n_j^1) \cdots Q(n_j^{l-1}) \tag{2}$$

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

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

(c) Let us assume that there are K symbols. Remember that for a valid dictionary we require all the paths in the tree to have atleast one word and prefix free means that the words should be leaves. Hence from every intermediate node there are K children and clearly $\frac{P(\text{child})}{P(\text{parent})} = p(k)$ where $p(k)$ is the probability of the symbol k. As a result

$$H_n = - \sum_{n \,:\, n \text{ is a child of } n'} \frac{P(n)}{P(n')} \log \frac{P(n)}{P(n')} = - \sum_k p(k) \log p(k) = H$$

Thus each $H_n = H$. Thus $H(\text{leaves}) = H \sum_{n \in I} P(n) = H\mathbb{E}[L]$.