ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 4 Information Theory and Coding
Solution 2 October 5, 2010, SG1 – 15:15pm-17:00

Problem 1 (Weak Law of Large Numbers)**.**

(a) (Markov's Inequality) Let the probability density function of X be $p_X(x)$. As X is non-negative and $a > 0$,

$$E[X] = \underbrace{\int_0^a x p_X(x) dx}_{\geq 0} + \int_a^\infty \underbrace{x}_{\geq a} p_X(x) dx \geq a \int_a^\infty p_X(x) dx = a \Pr\{X \geq a\}.$$

(b) (Chebyshev's Inequality) Set $X = (Y - \mu)^2$. Then, $E[X] = E[(Y-\mu)^2] = \sigma^2$. Using Markov's inequality,

$$\Pr\{|Y - \mu| \geq b\} = \Pr\{(Y-\mu)^2 \geq b^2\} = \Pr\{X \geq b^2\} \leq \frac{E[X]}{b^2} = \frac{\sigma^2}{b^2}.$$

(c) (Weak Law of Large Numbers) As $X_1, \cdots, X_n$ are IID, $E[\overline{X}_n] = E[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n E[X_i] = \mu$ and

$$\begin{aligned}
Var[\overline{X}_n] &= E[(\overline{X}_n - E[\overline{X}_n])^2] \\
&= E\left[\left(\overline{X}_n - \mu\right)^2\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)^2\right] \\
&= \frac{1}{n^2} E\left[\sum_{i=1}^n (X_i - \mu)^2 + \sum_{j=1}^n \sum_{k=1}^n (X_j - \mu)(X_k - \mu)\right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \underbrace{E[(X_i - \mu)^2]}_{=\sigma^2 \text{ (identical dist.)}} + \sum_{j=1}^n \sum_{k=1}^n \underbrace{E[(X_j - \mu)(X_k - \mu)]}_{=0 \text{ (independent)}} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Using Chebyshev's inequality for $\overline{X}_n$, we get

$$\Pr\{|\overline{X}_n - \mu| \geq \epsilon\} \leq \frac{Var[\overline{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}, \quad \forall \epsilon > 0.$$

Problem 2 (Jensen's Inequality)**.** When $n = 2$, $E[f(X)] = p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 f(x_1) + p_2 f(x_2)) = f(E[X])$, because f is convex. Assume Jensen's inequality is true for $n = k \geq 2$. If we show that it is true for $(k+1)$, we will have proven Jensen's inequality using mathematical induction. Let X take values $x_1, \cdots, x_{(k+1)}$ with probabilities $p_1, \cdots, p_{(k+1)}$, such that $\sum_{i=1}^{(k+1)} p_i = 1$. Let $p_i' = p_i/(\sum_{j=1}^{k} p_j)$ so that $\sum_{i=1}^{k} p_i' = 1$. From our assumption, we know that for $n = k$

$$E[f(X)] \geq f(E[X]) \Longleftrightarrow \sum_{i=1}^{k} f(x_i)p_i' \geq f\left(\sum_{i=1}^{k} x_i p_i'\right).$$

Now, for $n = k+1$

$$\begin{aligned}
E[f(X)] &= \sum_{i=1}^{k+1} f(x_i)p_i \\
&= \underbrace{\sum_{j=1}^{k} p_j}_{=(1-p_{(k+1)})} \sum_{i=1}^{k} f(x_i) \underbrace{\frac{p_i}{\sum_{j=1}^{k} p_j}}_{=p_i'} + f(x_{(k+1)})p_{(k+1)} \\
&= (1-p_{(k+1)}) \underbrace{\sum_{i=1}^{k} f(x_i)p_i'}_{\geq f(\sum_{i=1}^{k} x_i p_i')} + f(x_{(k+1)})p_{(k+1)} \\
&\geq (1-p_{(k+1)}) f\left(\sum_{i=1}^{k} x_i p_i'\right) + f(x_{(k+1)})p_{(k+1)} \qquad \text{(by assumption)} \\
&\geq f\left((1-p_{(k+1)}) \sum_{i=1}^{k} x_i p_i' + p_{(k+1)} x_{(k+1)}\right) \qquad \text{(because f is convex)} \\
&= f\left(\sum_{i=1}^{k+1} x_i p_i\right) = f(E[X]).
\end{aligned}$$

Problem 3 (Huffman Coding). (i) Given $Y = y_1$, the conditional probability distribution of X is

X	x_1	x_2	x_3	x_4	x_5	x_6
$\Pr\{X\|Y=y_1\}$	1/3	1/3	1/12	1/12	1/12	1/12

A possible binary Huffman tree for the above conditional distribution looks like Figure 1. From the tree, the expected codeword length W_1 of a binary Huffman code for X given $Y = y_1$ is $W_1 = 1/3 + 2/3 + 0 + 4/3 = 7/3$ bits.

(ii) Given $Y = y_2$, the conditional probability distribution of X is

X	x_1	x_2	x_3	x_4	x_5	x_6
$\Pr\{X\|Y=y_2\}$	1/6	1/6	1/6	1/6	1/6	1/6

A possible binary Huffman tree for the above conditional distribution looks like Figure 2. From the tree, the expected codeword length W_2 of a binary Huffman code for X given $Y = y_2$ is $W_2 = 0 + 2/3 + 6/3 = 8/3$ bits.

(iii) When Y is random, the probability distribution of X is

X	x_1	x_2	x_3	x_4	x_5	x_6
$\Pr\{X\}$	7/24	7/24	5/48	5/48	5/48	5/48

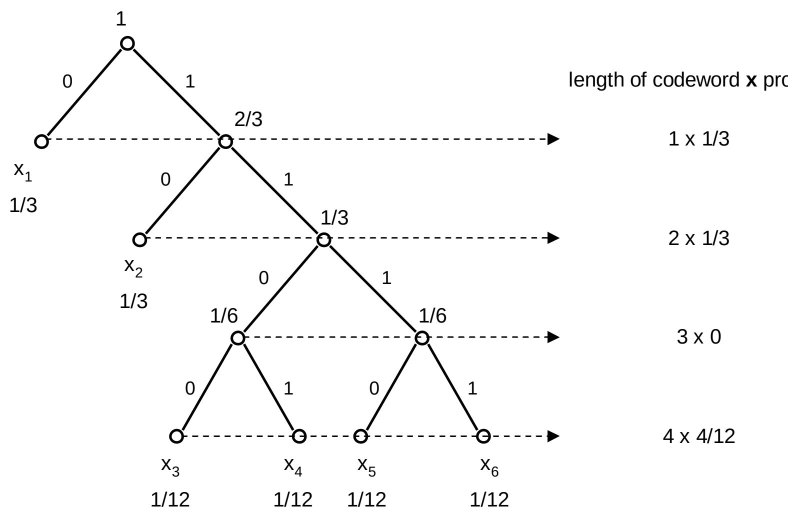

Figure 1: A possible binary Huffman code construction for Problem 3(i).

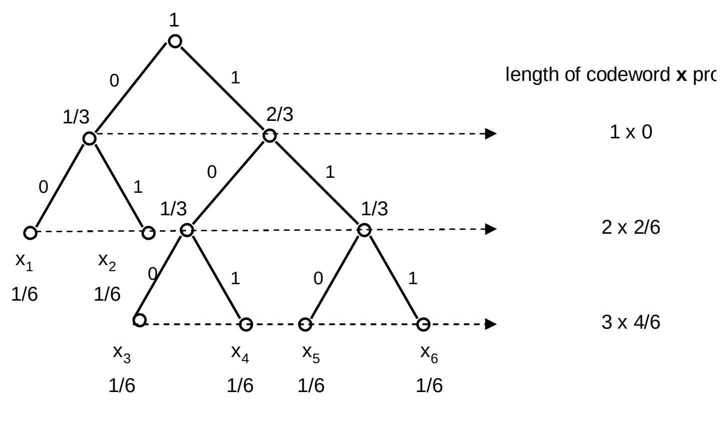

Figure 2: A possible binary Huffman code construction for Problem 3(ii).

A possible binary Huffman tree for the above distribution looks like Figure 3. From the tree, the expected codeword length W of a binary Huffman code for X is $W = 0 + 14/12 + 15/12 = 29/12$ bits. $W = 29/12 = 3/4 \times 7/3 + 1/4 \times 8/3 = \Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2$. So yes, $W = \Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2$ for this particular joint distribution.

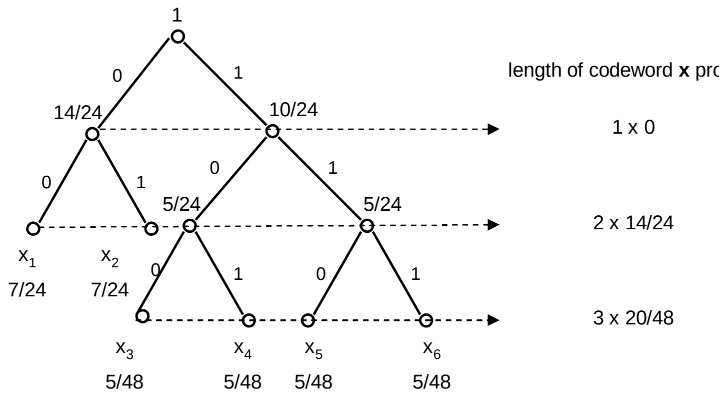

Figure 3: A possible binary Huffman code construction for Problem 3(iii).

(iv) No, it is not true in general. Consider the following joint distribution:

$Y \setminus X$	x_1	x_2	x_3
y_1	1/4	1/24	1/24
y_2	1/6	1/3	1/6

For this distribution, it can be verified that $W_1 = 5/4$ bits, $W_2 = 3/2$ bits, and $W = 19/12$ bits. However, $\Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2 = 1/3 \times 5/4 + 2/3 \times 3/2 = 17/12$ bits.

(v) Consider $Z = (X, Y)$ as a random variable with $\Pr\{Z = (x_i, y_j)\} = \Pr\{x_i, y_j\}$, $i = 1, \cdots, 6$, $j = 1, 2$. Then one possible Huffman tree looks like Figure 4. From the tree, the expected codeword length for Z is $0 + 1 + 0 + 4/3 + 5/6 = 19/6$ bits.

Problem 4 (Code Mismatch)**.** The expected codeword length is $L = \sum_{i=1}^{n} p_i l_i$. As $\log(1/q_i) \leq l_i \leq \log(1/q_i) + 1$, we get

$$\sum_{i=1}^{n} p_i \log \tfrac{1}{q_i} \leq L \leq \sum_{i=1}^{n} p_i \left(\log \tfrac{1}{q_i} + 1\right)$$

$$\underbrace{\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}}_{=D(p||q)} + \underbrace{\sum_{i=1}^{n} p_i \log \frac{1}{p_i}}_{=H(p)} \leq L \leq \underbrace{\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}}_{=D(p||q)} + \underbrace{\sum_{i=1}^{n} p_i \log \frac{1}{p_i}}_{=H(p)} + \underbrace{\sum_{i=1}^{n} p_i}_{=1}$$

$$H(p) + D(p||q) \leq L \leq H(p) + D(p||q) + 1.$$

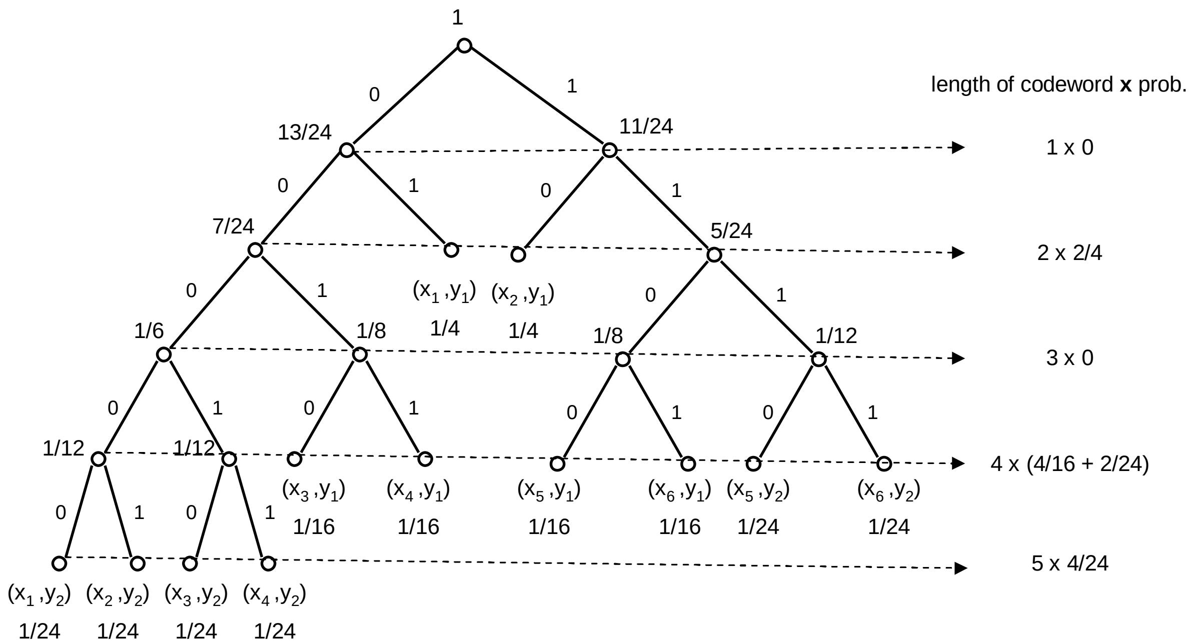

Figure 4: A possible binary Huffman code construction for Problem 3(v).