Homework 5 (2023) — SOLUTIONS

JAIST — School of Information Science — I232 Information Theory

1. *Huffman codes for a single source versus vector source.*

   (a) Consider a source code for a single random variable X which takes values from $\{1, 2, 3\}$:

$$p_{\mathsf{X}}(x) = \begin{cases} \frac{1}{3} & x = 1 \\ \frac{1}{3} & x = 2 \\ \frac{1}{3} & x = 3 \end{cases}$$

   Find a binary Huffman code for this source X. What is the expected length?

   Solution: A possible Huffman code is $C(1) = 0, C(2) = 10, C(3) = 11$. The expected length is $\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 2 = 1.6667$ bits .

   (b) Now consider an $n = 3$ vector random variable, $\mathbf{X} = \mathsf{X}_1\mathsf{X}_2\mathsf{X}_3$, which takes on values from $\{111, 112, \ldots, 333\}$. Assuming the X_i are independent, find the joint distribution:

$$p_{\mathbf{X}}(x_1, x_2, x_3) = p_{\mathsf{X}}(x_1)p_{\mathsf{X}}(x_2)p_{\mathsf{X}}(x_3).$$

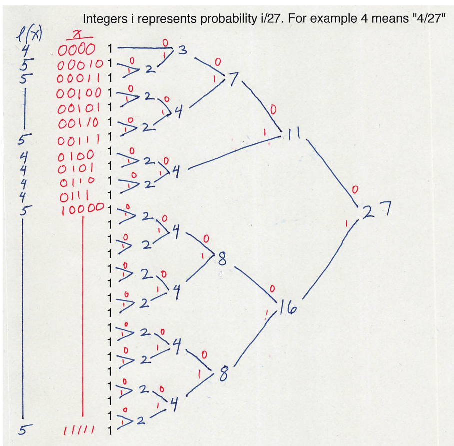
   Solution: Since $p_{\mathsf{X}}(x_i) = \frac{1}{3}$ by independence $p_{\mathbf{X}}(\mathbf{x})\mathbf{x} = \frac{1}{3}\frac{1}{3}\frac{1}{3} = \frac{1}{27} \approx 0.037037$.

   (c) Find a binary Huffman code for this vector source $\mathbf{X}$. What is the expected length? What is the expected length per source symbol (that is, the expected length divided by n).

   Solution: The tree is shown below, which includes the Huffman codewords $\mathbf{x}$ and their length $\ell(\mathbf{x})$. There are 5 codewords of length four and 22 codewords of length five, and each codeword has equal probability $1/27$. The average length and average length per source symbol are:

$$E[\ell] = \sum_{x \in \mathcal{X}} p_{\mathsf{X}}(x)\ell(x) = \frac{5}{27} \cdot 4 + \frac{22}{27} \cdot 5 = 4.8148 \text{ bits}$$

$$\frac{1}{n}E[\ell] = 1.6049 \text{ bits per symbol}$$



   (d) Which is better, single-symbol compression or vector compression? If we allow n to become large, what is the best possible compression in bits/symbol for this source?

Solution: Since 1.6059 bits/symbol is less than 1.6666 bits/symbol, vector compression is better.

The best possible compression rate is $H(\mathsf{X}) = \log 3 = 1.585$, which is even better, and can be achieved by asymptotically long vectors n.