Homework 4 (2023) — SOLUTIONS

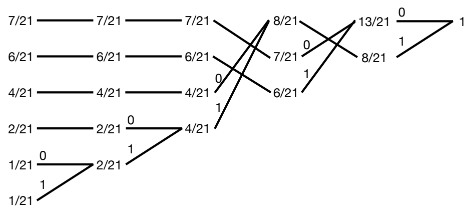
JAIST — School of Information Science — I232 Information Theory

1. Consider the following random variable distribution:

$$p_{\mathsf{X}}(x) = (\frac{1}{21}, \frac{1}{21}, \frac{2}{21}, \frac{4}{21}, \frac{6}{21}, \frac{7}{21}).$$

(a) Find a binary Huffman code

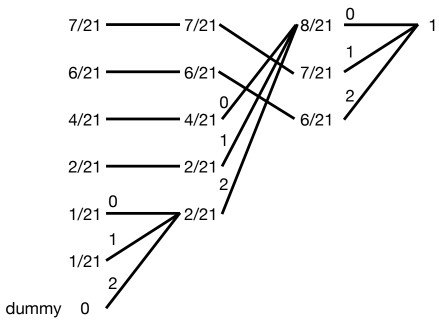
Solution: There are many possible codes. One is:



$p_{\mathsf{X}}(x)$	code
$\frac{7}{21}$	00
$\frac{6}{21}$	01
$\frac{4}{21}$	10
$\frac{2}{21}$	110
$\frac{1}{21}$	1110
$\frac{1}{21}$	1111

(b) Find a ternary Huffman code

Solution: Note that it is necessary to add a "dummy symbol" for the first step of the tree-generation. The number of dummy symbols can be calculated by $(1-|\mathcal{X}|) \mod (D-1) = -5 \mod 2 = 1$. There are many possible solutions. One is:



$p_{\mathsf{X}}(x)$	code
$\frac{7}{21}$	1
$\frac{6}{21}$	2
$\frac{4}{21}$	00
$\frac{2}{21}$	01
$\frac{1}{21}$	020
$\frac{1}{21}$	021

(c) Calculate $L = \sum_i p_{\mathsf{X}}(x)\ell_i$ for each case.

Solution:
(a) $L = 2\cdot\frac{7}{21} + 2\cdot\frac{6}{21} + 2\cdot\frac{4}{21} + 3\cdot\frac{2}{21} + 4\cdot\frac{1}{21} + 4\cdot\frac{1}{21} = \frac{16}{7} \approx 2.2857$
(b) $L = 1\cdot\frac{6}{21} + 1\cdot\frac{5}{21} + 2\cdot\frac{4}{21} + 2\cdot\frac{3}{21} + 3\cdot\frac{1}{21} + 3\cdot\frac{1}{21} = \frac{31}{21} \approx 1.4762$

2. KL divergence $D(p||q)$ is the cost of miscoding. Consider a source X distributed as p. The optimal code has expected length $H(\mathsf{X})$. If instead of the optimal code, we used the code optimal for q, then the expected length increases to $H(\mathsf{X}) + D(p||q)$. Thus, $D(p||q)$ is the cost of miscoding.

Let the random variable X have $\mathcal{X} = \{1, 2, 3, 4, 5\}$. Consider two distributions p_i and q_i on X:

Symbol	p_i	q_i
1	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{4}$	$\frac{1}{8}$
3	$\frac{1}{8}$	$\frac{1}{8}$
4	$\frac{1}{16}$	$\frac{1}{8}$
5	$\frac{1}{16}$	$\frac{1}{8}$

(a) Calculate $h(p)$, $h(q)$, $D(p||q)$ and $D(q||p)$.

(b) Find a Huffman code C_1 and its expected length $L(C_1)$ for source p.

(c) Find a Huffman code C_2 and its expected length $L(C_2)$ for source q.

(d) Show that $L(C_1)$ and $L(C_2)$ from the previous step satisfy the entropy bound.

(e) Now assume that we use code C_2 when the distribution is p. What is the average length of the codeword? By how much does it exceed the entropy p?

(f) What is the loss if we use code C_1 when the distribution is q?

Solution:

(a) $H(\mathsf{X}_p) = \sum_{i=1}^{5} p_i \log \frac{1}{p_i} = 1.875$. $H(\mathsf{X}_q) = \sum_{i=1}^{5} q_i \log \frac{1}{q_i} = 2$. $D(p||q) = D(q||p) = 0.125$

(b) We give an example of Huffman code C_1 from source p as

Symbol	$p_{\mathsf{X}}(x)$	code
1	$\frac{1}{2}$	0
2	$\frac{1}{4}$	10
3	$\frac{1}{8}$	110
4	$\frac{1}{16}$	1110
5	$\frac{1}{16}$	1111

The expected length is

$$\begin{aligned} L(C_1) &= \sum_i \ell_1(i) p_i \\ &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} = 1.875. \end{aligned}$$

(c) We give an example of Huffman code C_2 from source q as

Symbol	$p_{\mathsf{X}}(x)$	code
1	$\frac{1}{2}$	0
2	$\frac{1}{8}$	110
3	$\frac{1}{8}$	111
4	$\frac{1}{8}$	100
5	$\frac{1}{8}$	101

The expected length is

$$\begin{aligned} L(C_2) &= \sum_i \ell_2(i) q_i \\ &= 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} = 2. \end{aligned}$$

(d) Compute $H(p) = 1.875$ and $H(q) = 2$. The entropy bound is satisfied that $H(p) \le L(C_1) \le H(p)+1$ and $H(q) \le L(C_2) \le H(q)+1$. Since distribution p and q are 2-adic, $H(p) = L(C_1)$ and $H(q) = L(C_2)$ are achieved.

(e) Using C_2 with distribution p gives average length 2. It exceeds the entropy by 0.125 (which is exactly the divergence $D(p||q)$).

$$\begin{aligned} L(C_2, p) &= \sum_i \ell_2(i) p_i \\ &= 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{16} + 3 \cdot \frac{1}{16} = 2. \end{aligned}$$

(f) Using C_1 with distribution q gives average length 2.125. It exceeds the entropy by 0.125 (which is exactly the divergence $D(q||p)$).

3. Let source X be distributed according p_x for $x = 1, 2, \ldots, m$. C is a D-ary code for X with lengths ℓ_x for $x = 1, 2, \ldots, m$. Minimize $L(C)$ using Lagrange multipliers. That is, find $\ell_1, \ldots, \ell_m$ that minimize:

$$\min_{\ell_1,\ldots,\ell_m} \sum_{x=1}^{m} p_x \ell_x$$

subject to the restriction that the codes are a prefix code:

$$\sum_{x=1}^{m} D^{-\ell_x} \le 1.$$

To use Lagrange multipliers, ignore the restriction that ℓ_x are integers and assume ℓ_x are non-negative real numbers.

Solution: Minimize $L(C) = \sum_x p_x \ell_x$ over integers $\ell_1, \ell_2, \ldots$ subject to $\sum D^{-\ell_x} \le 1$. Hard to perform minimization over integers. Instead, treat $\ell_1, \ell_2, \ldots, \ell_m$ as real numbers. Assume equality on constraint.

Form Lagrangian:

$$J \;=\; \sum p_x \ell_x + \lambda\big(\sum D^{-\ell_x}\big)$$

Take the m partial derivatives with respect to ℓ_x:

$$\frac{\partial J}{\partial \ell_x} = p_x - \lambda D^{-\ell_x} \log_e D$$

set equal to 0:

$$D^{-\ell_x} \;=\; \frac{p_x}{\lambda \log_e D}$$

To find λ, substitute into the constraint

$$\begin{aligned} \sum D^{-\ell_x}\Big|_{D^{-\ell_x} = \frac{p_x}{\lambda \log_e D}} &= 1 \\ \sum \frac{p_x}{\lambda \log_e D} &= 1 \\ \lambda &= \frac{1}{\log_e D} \end{aligned}$$

Now find $D^{-\ell_x}$ using λ:

$$\begin{aligned} D^{-\ell_x} &= \frac{p_x}{\lambda \log_e D}\Big|_{\lambda = \frac{1}{\log_e D}} \\ D^{-\ell_x} &= p_x \text{ or} \\ \ell_x &= -\log_D p_x \end{aligned}$$

(Note: These non-integer ℓ_x give $L = \sum_{x\in\mathcal{X}} p_x \ell_x = -\sum_{x\in\mathcal{X}} p_x \log p_{\mathsf{X}}(x) = H(\mathsf{X})$. But since the lengths must be integers, we cannot achieve $L = H(\mathsf{X})$ unless p_x is D-adic.)