Lab 4 Report

Source Coding

Author: B08901049 Yuan-Chia Chang
Instructor: Professor Hao-Chung Cheng
TA: Yuan-Pon Chen
Created: 11/15/2021(Mon)
Last edited: 11/16/2021(Tue)
Contact: b08901049@ntu.edu.tw
Collaborator: B08901002 Chen-Han Lin, B08901209 Yu-Hsiang Lin

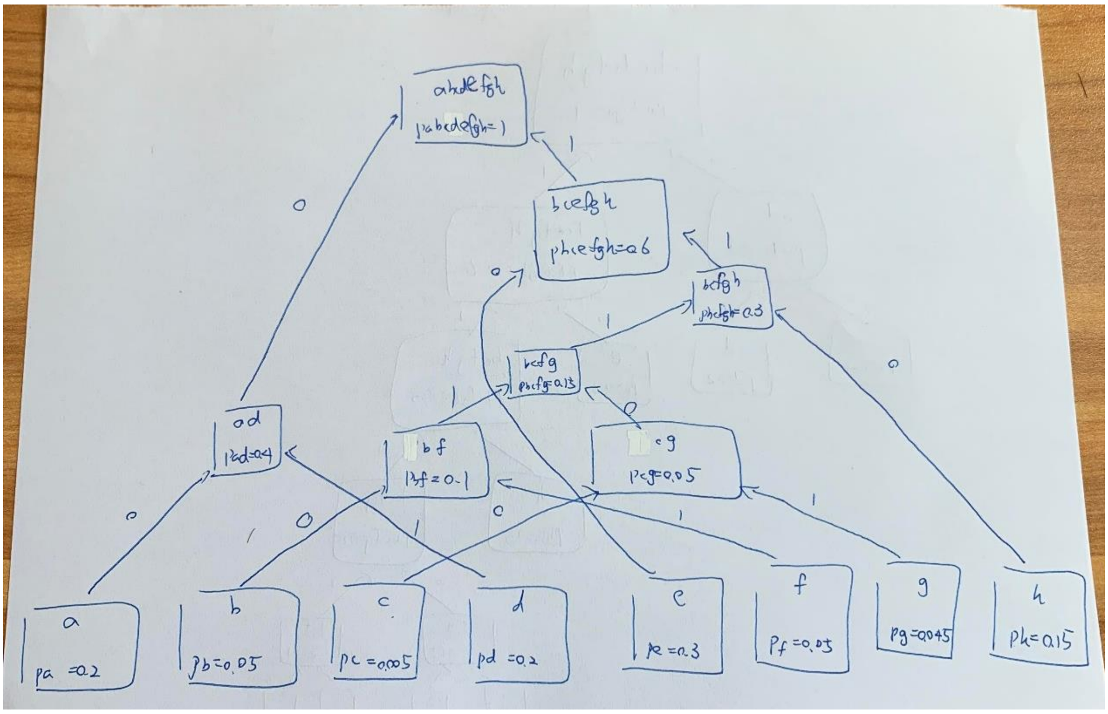
Q1.

(a)

$$\mathrm{H[X]} = -\sum_j p_j log p_j$$

$$= -(0.2 \times log(0.2) + 0.05 \times log(0.05) + 0.005 \times log(0.005) + 0.2 \times log(0.2) + 0.3 \times log(0.3) + 0.05 \times log(0.05) + 0.045 \times log(0.045) + 0.15 \times log(0.15)) \cong 2.5321$$

(b)



```
{'a'        }    {'00'     }
{'b'        }    {'11110'  }
{'c'        }    {'11100'  }
{'d'        }    {'01'     }
{'e'        }    {'10'     }
{'f'        }    {'11111'  }
{'g'        }    {'11101'  }
{'h'        }    {'110'    }
```

(c)

$$\sum_{j=1}^{M} 2^{-l(a_j)} = 2^{-2} + 2^{-5} + 2^{-5} + 2^{-2} + 2^{-2} + 2^{-5} + 2^{-5} + 2^{-3} = 1 \leq 1$$

Which holds Kraft inequality.

(d)

$\mathrm{E}[\mathrm{l}(a_j)] = \sum_{j=1}^{M} p_j l(a_j) = (0.2 \times 2 + 0.05 \times 5 + 0.005 \times 5 + 0.2 \times 2 + 0.3 \times 2 + 0.05 \times 5 + 0.045 \times 5 + 0.15 \times 3) = 2.6$, which is larger than $\mathrm{H}[\mathrm{X}] \cong 2.5321$.
Hence, Huffman Coding satisfies the source coding theorem.

(e)

1110100111000011110

(f)

gacab

(g)

By definition,

$$T_\varepsilon^n := \left\{ x^n \in \mathcal{X}^n : \left| \frac{-\log p_{X^n}(x^n)}{n} - H[X] \right| < \varepsilon \right\}$$

Also, due to the independence of $x^1, x^2, .., x^n$, $\frac{-log p_{X^n}(x^n)}{n} = \frac{-\sum_{i=1}^{n} log p_X(x^i)}{n}$

Hence, we want to find $2.43 = H[X] - \varepsilon < \frac{\sum_{i=1}^{n} -log p_X(x^i)}{n} < H[X] + \varepsilon = 2.63$

$$\text{for n } = 10, 24.3 < \sum_{i=1}^{10} -logp_X(x^i) < 26.3$$

$$-logp_X(a) = 2.3219, -logp_X(b) = 4.3219, -logp_X(c) = 7.6439,$$
$$-logp_X(d) = 2.3219, -logp_X(e) = 1.7370, -logp_X(f) = 4.3219,$$
$$-logp_X(h) = 4.4739, -logp_X(h) = 2.7370$$

Hence, we can find 10 symbols abedehedde，$\sum_{i=1}^{10} -logp_X(x^i) = 2.3219 + 1.7370 + 4.3219 + 2.3219 + 1.7370 + 4.4739 + 1.7370 + 2.3219 + 2.3219 + 1.7370 = 25.0314$, which is between 24.3 and 26.3. Hence abedehedde is a typical set of X with ε = 0.1 and n = 10.

Q2

(a)[1]

{'a' }	{[0.2000]}	{0×0 double}	{0×0 double}	{'00' }
{'b' }	{[0.0500]}	{0×0 double}	{0×0 double}	{'11110' }
{'c' }	{[0.0050]}	{0×0 double}	{0×0 double}	{'11100' }
{'d' }	{[0.2000]}	{0×0 double}	{0×0 double}	{'01' }
{'e' }	{[0.3000]}	{0×0 double}	{0×0 double}	{'10' }
{'f' }	{[0.0500]}	{0×0 double}	{0×0 double}	{'11111' }
{'g' }	{[0.0450]}	{0×0 double}	{0×0 double}	{'11101' }
{'h' }	{[0.1500]}	{0×0 double}	{0×0 double}	{'110' }
{'cg' }	{[0.0500]}	{[3]}	{[7]}	{'1110' }
{'bf' }	{[0.1000]}	{[2]}	{[6]}	{'1111' }
{'cgbf' }	{[0.1500]}	{[9]}	{[10]}	{'111' }
{'hcgbf' }	{[0.3000]}	{[8]}	{[11]}	{'11' }
{'ad' }	{[0.4000]}	{[1]}	{[4]}	{'0' }
{'ehcgbf' }	{[0.6000]}	{[5]}	{[12]}	{'1' }
{'adehcgbf'}	{[1]}	{[13]}	{[14]}	{0×0 char}

(b)

bin_seq: 1110100111000011110

which is identical with the result of 1e.

(c)

sym_seq: gacab

which is identical with the result of 1f.

Q3.

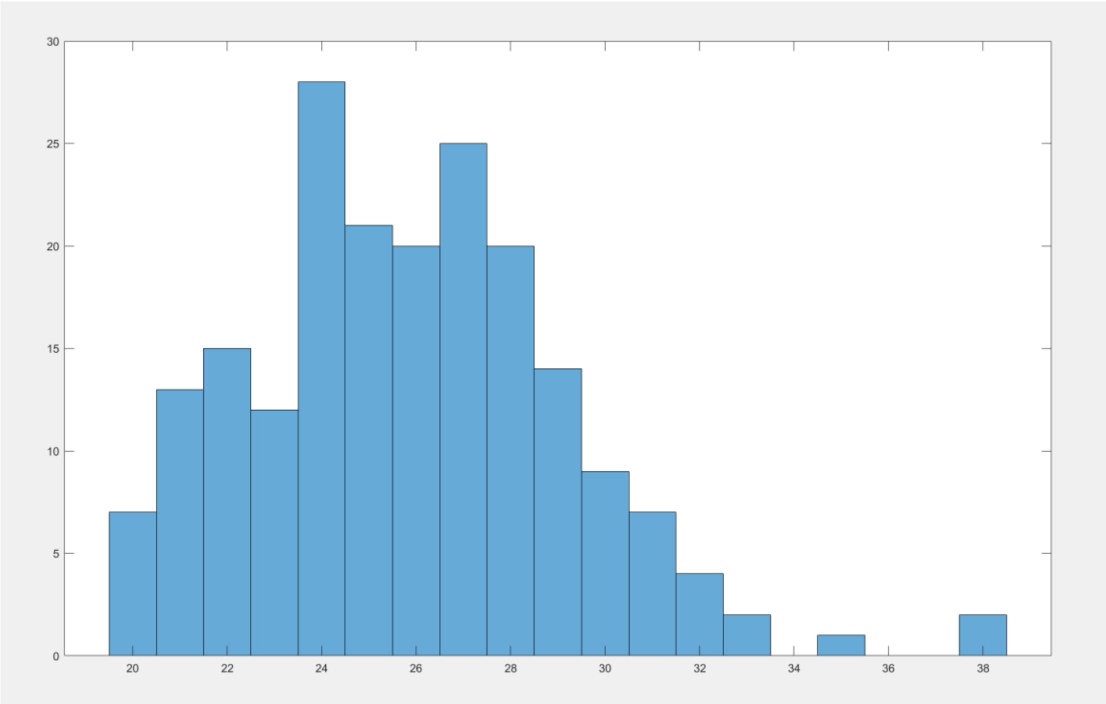
(a)

random sequence: deaheeaaed

random bit sequence: 011000110101000001001

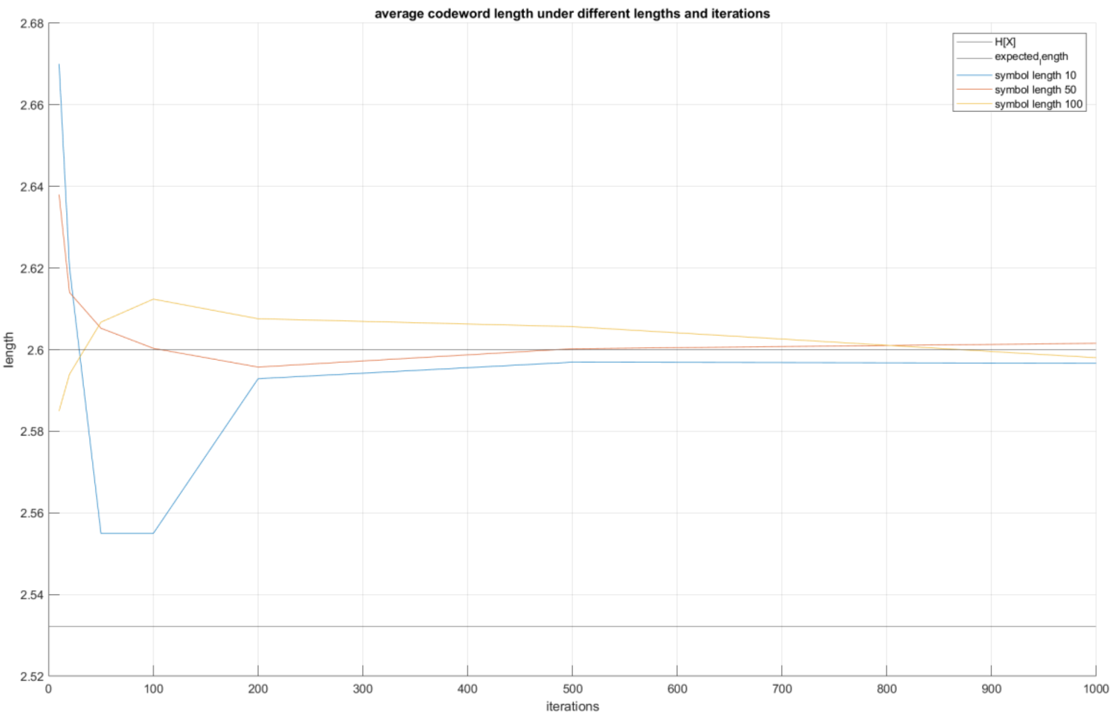
length of random bit sequence: 21

(b)



$$L_n(R) = 25.8450$$

(c)



(d)

For symbol length 10, 50, 100, the average codeword length converges to 2.6, which is the expected codeword length. However, symbol length 100 converges faster than symbol length 50, and symbol length 50 converges faster than symbol length 10.

For large iterations, the average codeword length is almost definitely larger than entropy according to source coding theorem.

Appendix

Code

https://github.com/yuanchiachang/CommLab/blob/main/Lab4/src

Reference

[1] https://www.cnblogs.com/klchang/p/13174608.html