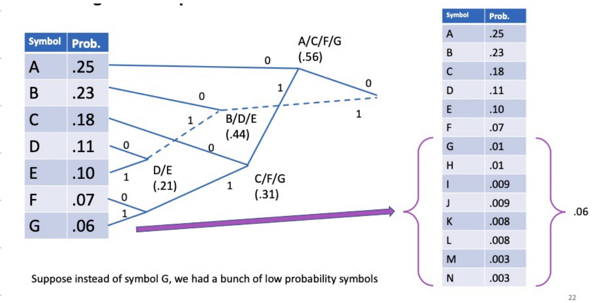
Homework 6

1. Truncated Huffman Coding.

Consider the alphabet w/ 14 possible symbols :

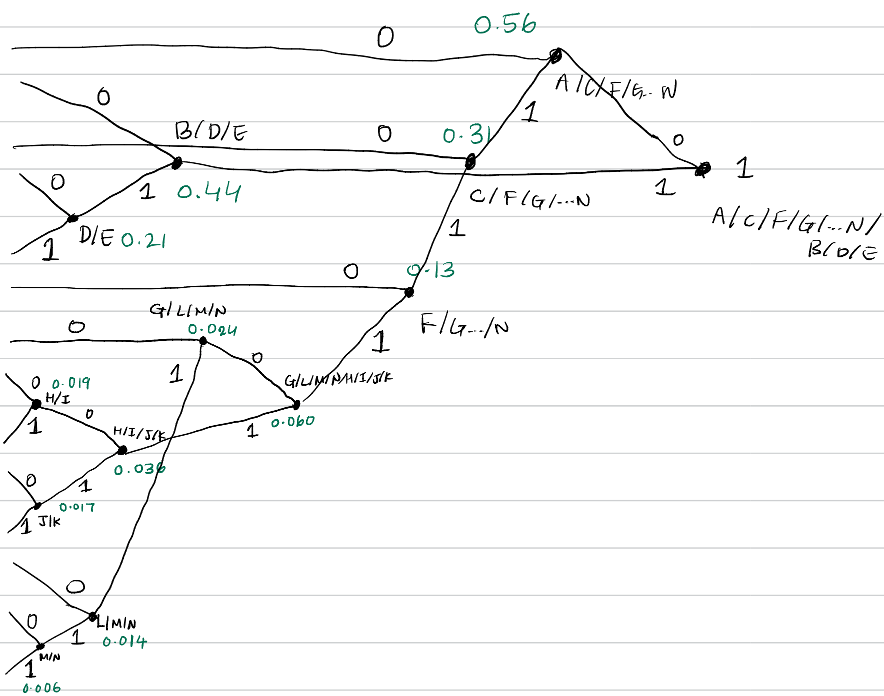


Symbol	Prob.
A	.25
B	.23
C	.18
D	.11
E	.10
F	.07
G	.06

Symbol	Prob.
A	.25
B	.23
C	.18
D	.11
E	.10
F	.07
G	.01
H	.01
I	.009
J	.009
K	.008
L	.008
M	.003
N	.003

(a)

Symbol	Prob
A	0.25
B	0.23
C	0.18
D	0.11
E	0.10
F	0.07
G	0.01
H	0.01
I	0.009
J	0.009
K	0.008
L	0.008
M	0.003
N	0.003



Codewords for each symbol

A	00	F	0110	K	0111111
B	10	G	011100	L	0111010
C	010	H	0111100	M	01110110
D	110	I	0111101	N	01110111
E	111	J	0111110		

(b) $$H(X) = \sum_{i=1}^{J} P(S_i) \log \frac{1}{P(S_i)}$$

$$= 0.25 \log\left(\frac{1}{0.25}\right) + 0.23 \log\left(\frac{1}{0.23}\right) + \cdots\cdots$$

$$+ 0.003 \log\left(\frac{1}{0.03}\right)$$

$$= 2.8009$$

```
In [1]: import math

        P = [0.25, 0.23, 0.18, 0.11, 0.10, 0.07, 0.01, 0.01, 0.009,0.009,0.008,0.008,0.003,0.003]

        def entropy(P):
            E = 0
            for p in P:
                E += p*math.log(1/p,2)
            return E
        entropy(P)

Out[1]: 2.800949958860649
```

length of $c(S_i)$

(c) $$EL(c) = \sum_{x \in A} P(x)\, l(x)$$

$$= 2.8260$$ ← Expected length

```
In [4]: P = [0.25, 0.23, 0.18, 0.11, 0.10, 0.07, 0.01, 0.01, 0.009,0.009,0.008,0.008,0.003,0.003]
        code = ['00', '10', '010', '110', '111', '0110', '011100', '0111100', '0111101', '0111110',
                '0111111', '0111010', '01110110', '01110111']
        L = [len(i) for i in code]

        E = 0
        for p,L in zip(P,L):
            E += p*L
        E

Out[4]: 2.8260000000000005
```

Efficiency of full Huffman code is

$$\frac{H}{L} = \frac{2.8009}{2.826} \approx 0.9911$$

Expected length of the truncated Huffman code is

$$EL(C_{trunc}) = \sum_{x \in A_{trunc}} p(x)\, l(x)$$

$$\approx 2.83$$

```
In [6]: P = [0.25, 0.23, 0.18, 0.11, 0.10, 0.07, 0.06] # 0.01, 0.01, 0.009,0.009,0.008,0.008,0.003,0.003]
        code = ['00', '10', '010', '110', '111', '0110', '0111000'] #, '011100', '0111100', '0111101', '0111110',
        #         '0111111', '0111010', '01110110', '01110111']
        L = [len(i) for i in code]

        E = 0
        for p,L in zip(P,L):
            E += p*L
        E

Out[6]: 2.83
```

Efficiency of truncated Huffman code is

$$\frac{2.8009}{2.83} \approx 0.9897$$

2. Arithmetic coding.

Four symbol source $\{a, b, c, d\}$

$p(S_i) = \{0.1, 0.4, 0.3, 0.2\}$

$cdf = \{0, 0.1, 0.5, 0.8, 1\}$

Sequence: bba

$a \in [0, 0.1)$

$b \in [0.1, 0.5)$

First symbol $b \rightarrow [0.1, 0.5)$

left of $1/2$

output 0

rescale $[0.2, 1)$

Next, symbol $b \rightarrow$ sub interval =

$0.2 + 0.1 \times 0.8 = 0.28$

$0.28 + 0.4 \times 0.8 = 0.6$

No output $= [0.28, 0.6)$

Next symbol $a \rightarrow$ New interval

$[0.28, 0.28 + (0.1)(0.6 - 0.28))$

$= [0.28, 0.312)$

endpoint left of $1/2$

output 0

rescale $[0.56, 0.624)$

right of 1/2

output 1

rescale $[0.12, 0.248)$

left of 1/2

output 0

rescale $[0.24, 0.496)$

left of 0.5

output 0

rescale $[0.48, 0.992)$

straddles 0.5

no output

3. $$G(u) = \frac{1}{\sqrt{2N}} \sum_{k=0}^{2N-1} g(k)\, e^{-j2\pi uk/2N}$$

$$G(u) = \frac{1}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k)\, e^{-j2\pi uk/2N} + \frac{1}{\sqrt{2N}} \sum_{k=N}^{2N-1} f(\overbrace{2N-1-k}^{\substack{k \\ = \\ z}})\, e^{-j2\pi u \overbrace{k}^{\substack{k \\ \uparrow \\ 2N-1-z}}/2N}$$

Change of variables
$z = 2N-1-k$
$k = 2N-1-z$

$$= \frac{1}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k)\, e^{-j(2\pi uk)/2N} + \frac{1}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k)\, e^{-j2\pi u(2N-1-k)/2N}$$

$$= \frac{1}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k) \left[e^{-j\frac{2\pi uk}{2N}} + e^{-j2\pi u\cancel{\frac{2N}{2N}} + j2\pi u(k+1)/2N} \right]$$

$$= \frac{1}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k)\, 2\cos\left[\frac{u\pi(2k+1)}{2N}\right] e^{ju\pi/2N}$$

$$= \frac{2}{\sqrt{2N}} \sum_{k=0}^{N-1} f(k) \cos\left(\frac{u\pi(2k+1)}{2N}\right) e^{ju\pi/2N}$$

$\therefore$ $C(u)$ and $G(u)$ are related by

$$C(u) = \frac{\alpha(u)}{2/\sqrt{2N}}\, G(u)\, e^{-ju\pi/2N}$$