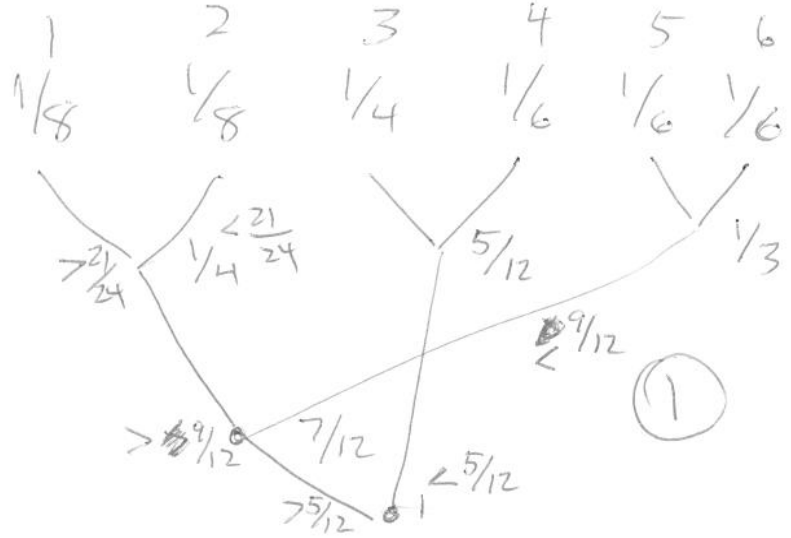
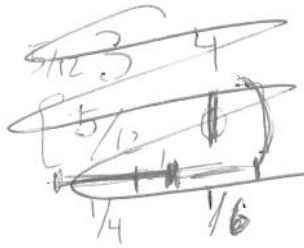


Optimal # of tests to perform to test which interval we are in

1



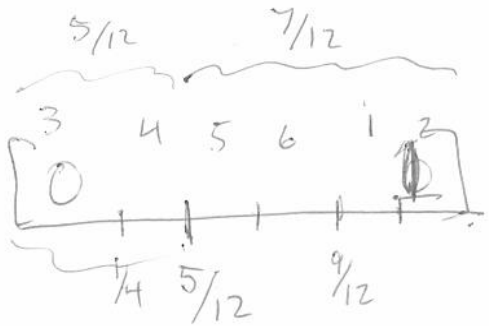
~~if w > 5/12 then~~

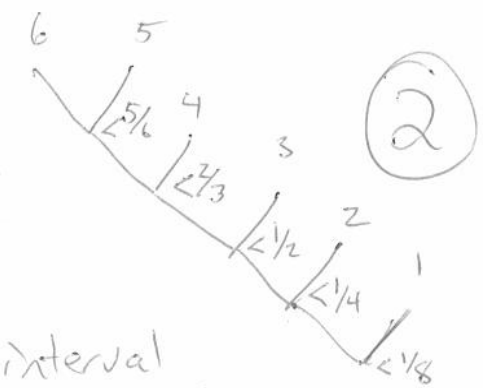
if w > 8/12 then ans = 4

~~if~~ else ans = 3

~~else~~

~~if~~





2

```
if  w < 5/12  then
   if w < 1/4  then ans = 3
    else            ans = 4
else if w > 9/12 then
   if w > 21/24 then ans = 1
    else            ans = 2
else  % w > 5/12 but < 9/12   comment
   if w > 7/12 then ans = 6
    else            ans = 5
```

another way of testing which interval we are in, but this has expected # of test higher than Huffman.

For ~~xx~~ tree ② Expected # of tests

	1	2	3	4	5	6
# of tests	1	2	3	4	5	5
prob	1/8	1/8	1/4	1/6	1/6	1/6

$$E(\text{tests}) = 1/8 \cdot 1 + 1/8 \cdot 2 + 1/4 \cdot 3 + 1/6 \cdot 4 + 1/6 \cdot 5 + 1/6 \cdot 5$$

For for tree ① Expected # of tests

	1	2	3	4	5	6
# of tests	3	3	2	2	3	3
prob	1/8	1/8	1/4	1/6	1/6	1/6

$$E(\text{tests}) = 3 \cdot 1/8 + 3 \cdot 1/8 + 2 \cdot 1/4 + 2 \cdot 1/6 + 3 \cdot 1/6 + 3 \cdot 1/6$$

Expected Code Length

Theorem 2 *The best possible expected code length (bits per letter) is*

$$\boxed{H = \sum_{i=1}^{n} p_i \log_2 1/p_i}$$

Proof.

$$\text{Letter frequenices } N_1, N_2, \ldots, N_k \ (N = \textstyle\sum_{i=1}^{k} N_i)$$

$$\text{Code lengths } h_1, h_2, \ldots, h_k \text{ (from a binary tree)}$$

$$p_i = N_i/N \text{ and } q_i = 1/2^{h_i}$$

(Average bits per letter) · (# of letters) = File length

$$\begin{aligned} &= \sum_{i=1}^{k} N_i h_i \\ &= \sum_{i=1}^{k} N_i \log_2 2^{h_i} \\ &= N \sum_{i=1}^{k} p_i \log_2 1/q_i \\ &\geq N \sum_{i=1}^{k} p_i \log_2 1/p_i = NH \end{aligned}$$

The tree from heights has code length $\leq H+1$ bits per letter

tree from heights is a code for which $h_i = \lceil \log_2 1/p_i \rceil$

$$\text{ECL of tree from heights} = \sum_{i=1}^{n} N_i \cdot h_i$$

$$\cancel{= \sum_{i=1}^{n} N_i \log_2 2^{h_i}}$$

$$= \sum_{i=1}^{n} N_i \cancel{\log_2} 2^{\lceil \log_2 1/p_i \rceil}$$

$$\leq \sum_{i=1}^{n} N_i \cancel{\log_2} 2^{\log_2 1/p_i + 1}$$

$$= \sum_{i=1}^{n} N_i (\log_2 1/p_i + 1)$$

$$= \sum_{i=1}^{n} N \cdot p_i \log_2 1/p_i + \sum_{i=1}^{n} N_i$$

$$= N \cdot \underbrace{\sum_{i=1}^{n} p_i \log_2 1/p_i}_{H} + N$$

$$= N \cdot (H+1)$$

$$ECL \leq N \cdot (H+1)$$

Homophonic – What is unicity distance for key length 20?

	A B C	. . . .	I/J	. . N .	. S T .		Y Z
S					25 1 2 3	. . . . .	
T	.				49 50 26 27 28	. . . . .	
A	51 52 53	. . . . .					74 75
N				76 77 78	. . . . .		

Key length 20 how many keys are there?

25^{20} possible $\qquad H(K) = \log_2 25^{20}$

Cyphertexts of length N assuming all alphabets are equally likely: $(25 \cdot 20)^N \qquad H(C) = \log_2(500^N)$

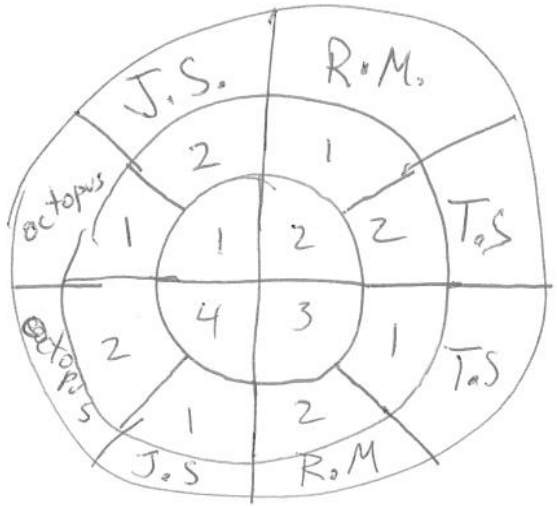
Plaintexts of length N (with English redundancy).

$$H(M) = N \cdot 1.5$$

$$H(C) = H(M) + H(K)$$

$$N \cdot \log_2 500 = N \cdot 1.5 + 20 \cdot \log_2 25$$

$$N = \frac{20 \cdot \log_2 25}{\log_2 500 - 1.5} = 12.4$$



inner = K
Middle = M
Outer = C