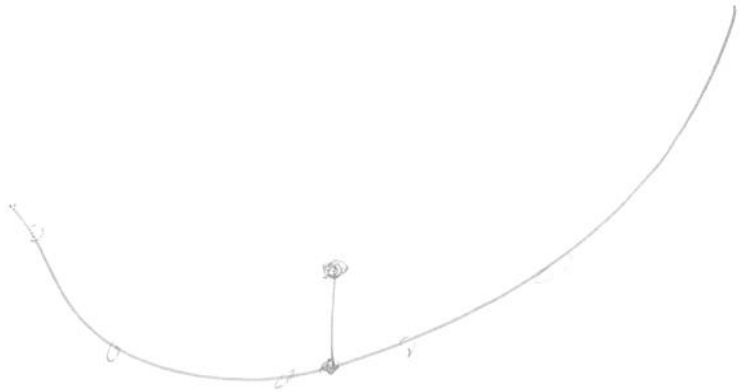
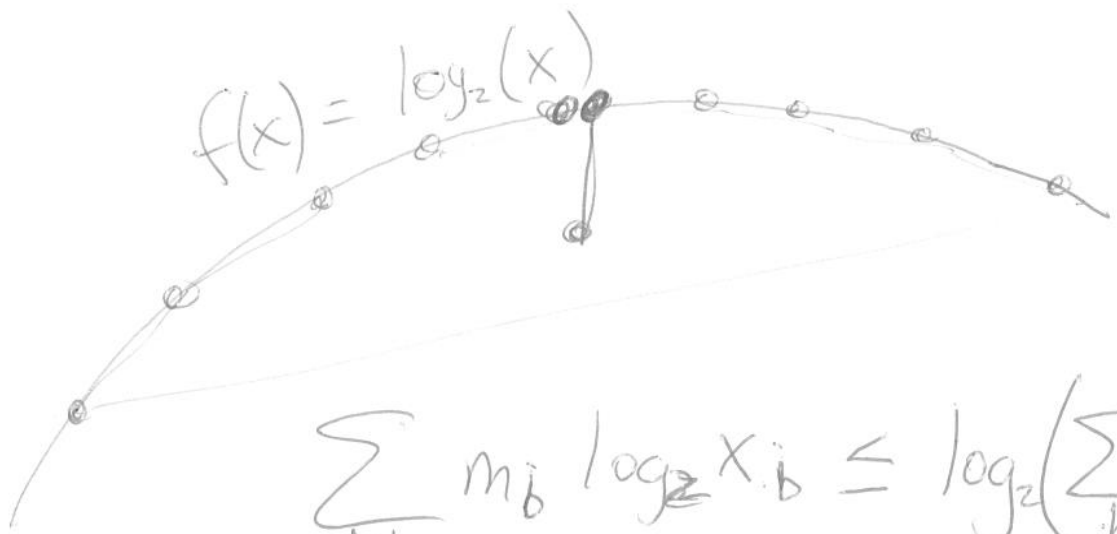




$$\sum_{b} m_b \log_2 x_b \leq \log_2\left(\sum_{b} m_b x_b\right)$$

Since for a given a, the conditional probabilities $P[Y = b|X = a]$ add up to 1, we can use the convex function inequality

$$\sum_b m_b \log_2 x_b \leq \log_2 \left(\sum_b m_b x_b \right)$$

For appropriate choices of m_b and x_b we have:

$$\begin{aligned}
\sum_b P[Y = b|X = a] \log_2 \frac{1}{P[X = a|Y = b]} &\leq \log_2 \left(\sum_b P[Y = b|X = a] \frac{1}{P[X = a|Y = b]} \right) \\
&= \log_2 \left(\sum_b \frac{\cancel{P[X = a, Y = b]}}{P[X = a]} \times \frac{P[Y = b]}{\cancel{P[X = a, Y = b]}} \right) \\
&= \log_2 \left(\sum_b \frac{P[Y = b]}{P[X = a]} \right) \\
&= \log_2 \frac{1}{P[X = a]}
\end{aligned}$$

Therefore

$$\begin{aligned}
H(X|Y) &= \sum_a P[X = a] \sum_b P[Y = b|X = a] \log_2 \frac{1}{P[X = a|Y = b]} \\
&\leq \sum_a P[X = a] \log_2 \frac{1}{P[X = a]} = H(X).
\end{aligned}$$

$\cancel{H(Y|X) \le H(X)}$

Theorem 2 *For any two random variables X and Y we always have*

$$H(X|Y) \le H(X) \tag{1}$$

and equality holds if and only if X and Y are independent.

$P[X=a|Y=b] = P[X=a]$

$P[X=a \& Y=b] = P[X=a]P[Y=b]$

Proof. From our definitions we get

$$\begin{aligned}
H(X|Y) &= \sum_b P[Y=b] H(X|Y=b) \\
&= \sum_b P[Y=b] \sum_a \cancel{P[X=a|Y=b]} \log_2 \frac{1}{\cancel{P[X=a|Y=b]}} \\
&= \sum_b \cancel{P[Y=b]} \sum_a \frac{P[X=a, Y=b]}{\cancel{P[Y=b]}} \log_2 \frac{1}{P[X=a|Y=b]} \\
&= \sum_b \sum_a \frac{P[X=a, Y=b]}{P[X=a]} \log_2 \frac{1}{P[X=a|Y=b]} \\
&= \sum_a P[X=a] \sum_b P[Y=b|X=a] \log_2 \frac{1}{P[X=a|Y=b]} \qquad (2)
\end{aligned}$$

$H(X)$

$P[X=a]$

$P[X=a]$

Assume $P[X=a|Y=b] = P[X=a]$

m_i

x_i

$\sum_b m_i \log_2 x_i$

$\sum_b \left(\cancel{m_i \log_2 m_i} \times \log_2 \left(\sum_b m_i x_i \right) \right)$

because K & M are independent.

$$\begin{aligned} H(K) + H(M) &= H(K, M) \\ &= H(K, C) = H(C, K) \\ &= H(C) + H(K|C) \end{aligned}$$

For a ciphertext only attack

$$\boxed{H(K|C) = H(K) + H(M) - H(C)}$$

Unicity Distance

Definition The *unicity distance* is the smallest number of characters in the ciphertext that uniquely determines the plaintext.

Since we assume that the ciphertext uniquely determines the plaintext, the key must also be determined. Therefore

$$\boxed{H(K|C) = 0}$$

or in other words

$$\boxed{H(K) = H(C) - H(M)}$$

First estimate of monoalphabetic

$$H(K) = \log_2(26!)$$

$$H(M) = 1.2N \quad = \text{entropy of Eng} \cdot N$$

$$H(C) = 4.7N = \log_2 26^N$$

At unicity distance we have

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

$$\log_2 26! = 4.7N - 1.2N$$

$$88.3 = 3.5N$$

$$N \cong 26$$

25....

Hill 2x2 mod 29.

Key $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

# of keys $= 29^4$

$$H(K) = \log_2 29^4 = 4 \cdot \log_2 29 \cong 19.4$$

$$H(C) = \log_2 29^N = N \log_2 29 = 4.9N$$

$$H(M) = 3.2N$$

$$19.4 = 4.9N - 3.2N$$

$$N \cong 11$$

Playfair

Keys

		D		
C			B	H
		G	I/J	
	A		E	
		F		

$$H(K) = \cancel{\log_2(25!)} = \log_2(24!)$$

$$H(M) = \cancel{2N} = 1.2N$$

$$H(C) = \log_2(25^N) = \log_2(25)N$$

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

$$\cancel{83.7} = 4.6N - \cancel{2N} = \cancel{2.6N}$$

$$79 = \cancel{32N} \quad \cancel{1.2N} \quad \cancel{3.4N}$$

$$2N \quad 2.6N$$

$$N \cong \cancel{32}$$

$$N \cong \cancel{25} \ \cancel{23}$$

$$\cong 30$$

Answer is somewhere in between 23 and 30

G	H	I/J	K	L
M	N	A	B	O
P	Q	C	D	R
S	T	E	F	V
V	W	X	Y	Z

$\cong$

I/J	K	L	G	H
A	B	O	M	N
C	D	R	P	Q
E	F	V	S	T
X	Y	Z	V	W

Encode AF $\longrightarrow$ BE

recal # of keys $= \frac{25!}{25} = 24!$

Expected code length for tree from heights:

If no branches are trimmed

$$ECL = \sum_{\alpha=A}^{Z} p_\alpha \cdot \underbrace{\lceil \log_2(1/p_\alpha) \rceil}_{\text{height at which each letter is encoded}}$$

If branches are trimmed (some letters are moved down)

$$ECL \leq \sum_{\alpha=A}^{Z} p_\alpha \lceil \log_2(1/p_\alpha) \rceil$$

$$\leq \sum_{\alpha=A}^{Z} p_\alpha \left(\log_2(1/p_\alpha) + 1 \right)$$

$$= \sum_{\alpha=A}^{Z} p_\alpha \log_2(1/p_\alpha) + \underbrace{\sum_{\alpha=A}^{Z} p_\alpha}_{1}$$

$$= H(\text{single letter stats}) + 1$$

Representing English letters as sequences of bits
ECL = expected code length

a	8.167%
b	1.492%
c	2.782%
d	4.253%
e	12.702%
f	2.228%
g	2.015%
h	6.094%
i	6.966%
j	0.153%
k	0.772%
l	4.025%
m	2.406%
n	6.749%
o	7.507%
p	1.929%
q	0.095%
r	5.987%
s	6.327%
t	9.056%
u	2.758%
v	0.978%
w	2.360%
x	0.150%
y	1.974%
z	0.074%

a=1
b=01
c=001
d=0001
e=00001
f=000001
g=0000001
h=00000001
i=000000001
j=0000000001
...

ECL ≈ 11.65 bits per character

EELS MAKE GOOD PETS

0000100001 0000 ... 01 000000001

e t a o i n s h r d l c u m w f g y p b v k j x q z

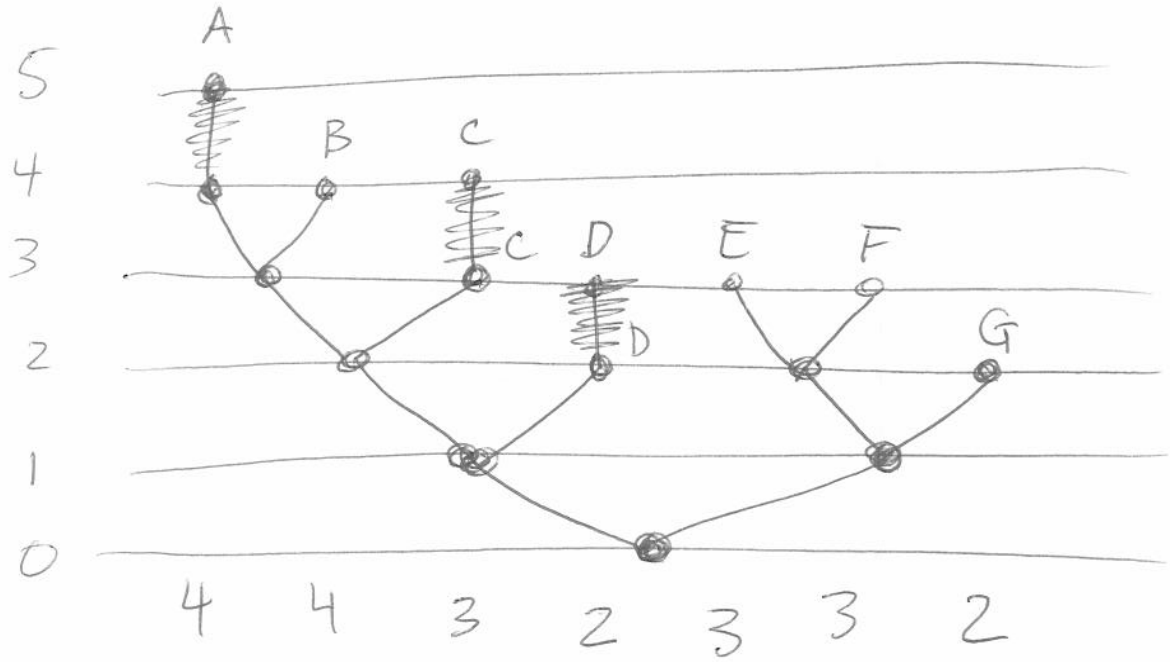
e=1
t=01
a=001
o=0001
i=00001
n=000001
s=0000001
h=00000001
r=000000001
d=0000000001
...

ECL ≈ 7.53 bits per character

Entropy of English using single letter stats $= \sum_{\alpha=A}^{Z} p_\alpha \log_2(1/p_\alpha) \approx 4.16$

Tree from heights

	A	B	C	D	E	F	G
	2	4	6	10	13	13	16
$\lceil \log_2(1/p) \rceil$	5	4	4	3	3	3	2



Expected code length =

$$\frac{2\cdot 4 + 4\cdot 4 + 3\cdot 6 + 2\cdot 10 + 3\cdot 13 + 3\cdot 13 + 2\cdot 16}{64}$$

$$2.7 \cong \frac{172}{64} = 62 + 39 + 39 + 32$$

Huffman Code

Begin with a text file with the following frequencies

letter	A	B	C	D	E	F	G
frequency	2	4	6	10	13	13	16
code length	5	5	4	3	2	2	2

-

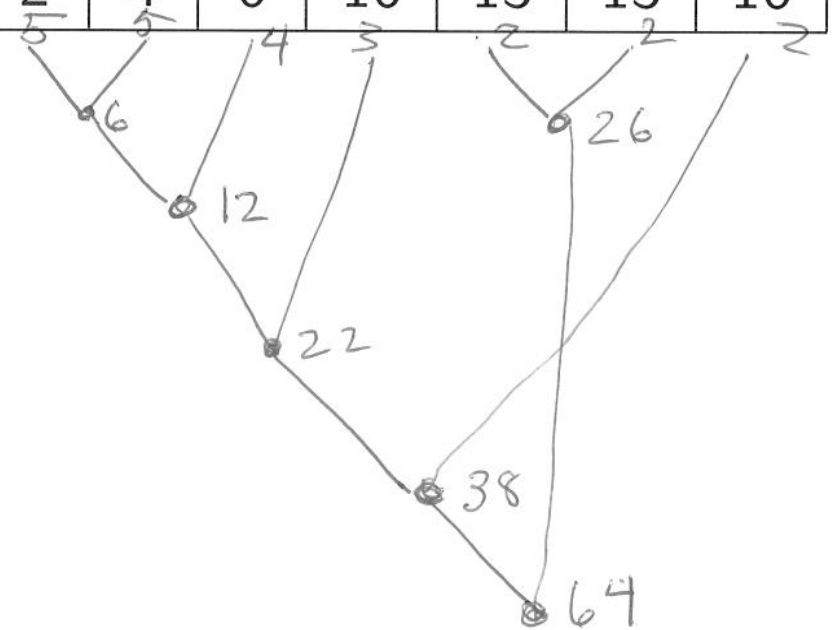
$$\text{average bits per letter} = (5 \cdot 2 + 5 \cdot 4 + 4 \cdot 6 + 3 \cdot 10 + 2 \cdot 13 + 2 \cdot 13 + 2 \cdot 16)/64 = \frac{168}{64} = 2.625$$

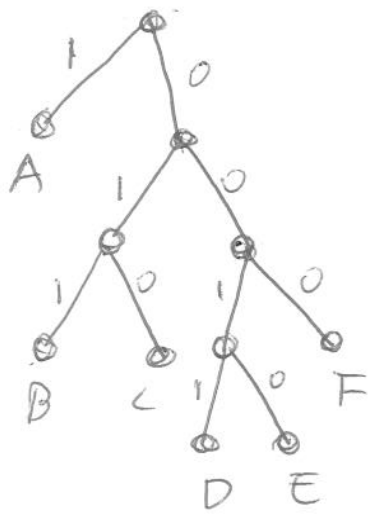
Entropy $\cong 2.579$

Huffman Code

Begin with a text file with the following frequencies

letter	A	B	C	D	E	F	G
frequency	2	4	6	10	13	13	16





BADEF

011100110010000

1111111101010010110

AAAAAAAACAFAB