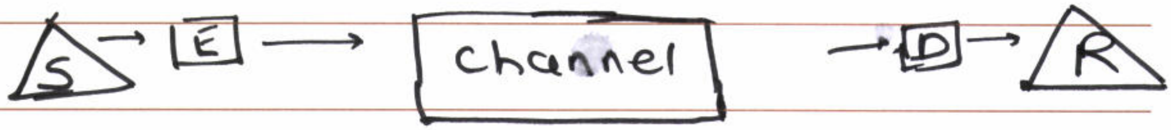
Information Theory

Claude Shannon



1) how do you efficiently transmit data from $S \to R$

source coding

2) how do you reliably transmit dat from $S \to R$

channel coding

source $\to$ vocabulary

$S \to \{ s_1 \quad s_2 \quad s_3 \quad s_4 \}$

s_1	70	1	70
s_2	5	001	15
s_3	20	01	40
s_4	5	000	15
			140

1.4 bits/symbol

$$S \rightarrow \{ s_1 \quad s_2 \quad s_3 \quad \ldots \quad s_N \}$$
$$l_1 \quad l_2 \quad l_3 \quad \ldots \quad l_N$$

$S \rightarrow$ a message of length M in time T

In $M \rightarrow m_1 \quad m_2 \quad m_3 \quad \ldots \quad m_N$

$$M = m_1 + m_2 + \cdots + m_N = \sum m_i$$

$$\text{\# of bits in } M = m_1 l_1 + m_2 l_2 + \cdots + m_n l_n$$
$$L = \sum m_i l_i$$

$S \rightarrow \{s_1, s_2, s_3, s_4\}$

$S \rightarrow s_1\ s_1\ s_2\ s_1\ s_4\ s_3\ s_1 \ldots$

s_1	70	00 →	140
s_2	5	01	10
s_3	20	10	40
s_4	5	11	10
			200

2 bits / symbol

s_1	70	0	70
s_2	5	01	10
s_3	20	1	20
s_4	5	10	10
			110

1.1 bits / symbol

$s_1 s_3 \rightarrow 01$

$s_2 \quad 01$

t = avg bits per symbol

$$L = \sum m_i l_i$$

$$t = \frac{L}{M} = \frac{\sum m_i l_i}{M}$$

$$t = \sum \left(\frac{m_i}{M}\right) l_i$$

$$= \sum P_i l_i$$

minimize t

$$H = \sum P_i \log 1/P_i$$

$$= - \sum P_i \log P_i$$

Entropy

Lossless	Lossy
no loss	you don't get original signal
VBR	CBR
! guaranteed	guaranteed

- Lossless
  - Removing repetition
    - run length
    - repitition supression
  - Pattern based
    - ~~LZM~~ LZW
  - Statistical
    - Huffman
    - Arithmetic
    - Shannon Fano
    - Goulomb
- Lossy
  - Signal Importance
    - quantization
    - filtering
    - subsampling
  - Prediction based
    - DPCM
    - motion compensation
  - Frequency domain
    - Fourier
    - DCT
    - DST
    - wavelet
    - KLT
    - subband
  - Hybrid
    - JPEG
    - MPEG
    - ITU

Runlength
$S = \{a, b\}$

aaaaabbbaaabbbb ...
a5b3a3b4 ....

LZW $S = \{a, b\}$

a b b a a b b a a b a b ...

$\underbrace{a}_{0}\ \underbrace{b}_{1}\ \underbrace{b}_{1}\ \underbrace{a}_{0}\ \underbrace{a\ b}_{2}\ \underbrace{b\ a}_{4}$

a	0	ba	4
b	1	aa	5
ab	2	abb	6
bb	3	baa	7

Huffman coding

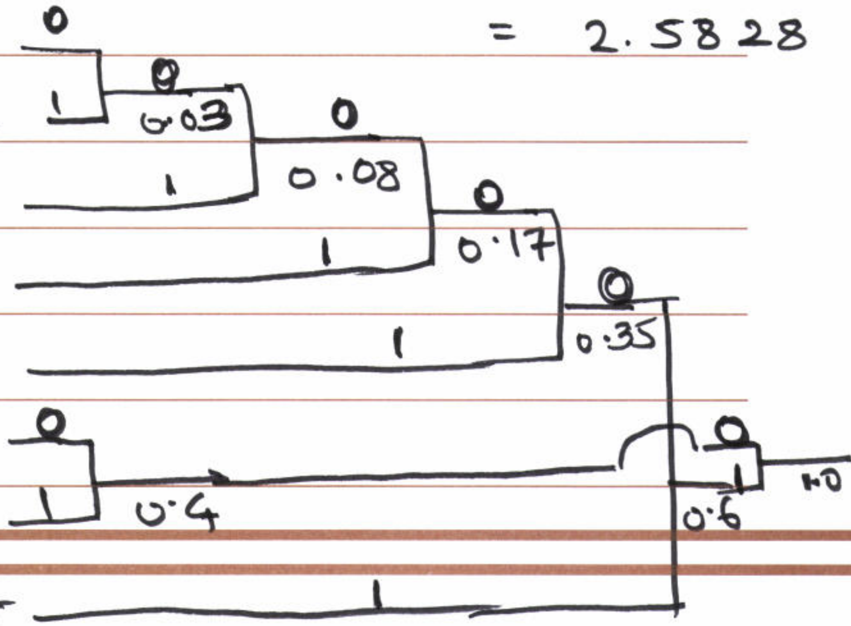
$S = \{a, b, c, d, e, f, g, h\}$

$\sum p_i l_i = 3$

$H = -\sum p_i \log p_i$

$= 2.5828$

00000	a	0·01
00001	b	0·02
0001	c	0·05
001	d	0·09
01	e	0·18
00	f	0·2
01	g	0·2
11	h	0·25



$\sum p_i l_i = 2.63$

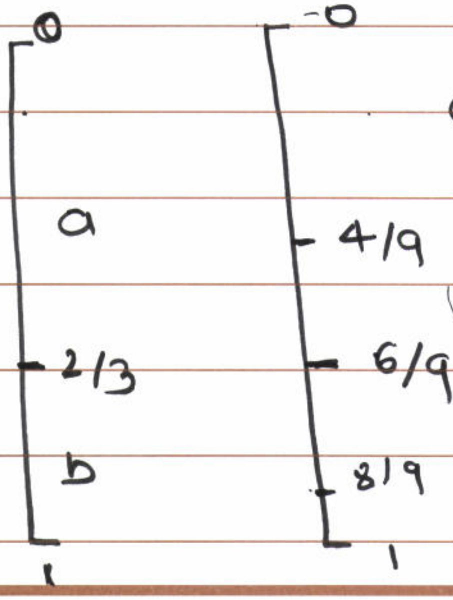
$$\text{efficiency} = \frac{H}{\text{avg code length}} = \frac{2.5828}{2.63}$$

Arithmetic

$S = \{a, b\}$

$\quad 2/3 \quad 1/3$

$($ $]$



aa	$\frac{1}{4}$	01
ab		1
ba		11
bb		1111

$\ldots 2^3 \quad 2^2 \quad 2^1 \quad 2^0 \quad 2^{-1} \quad 2^{-2} \ldots$

.

Lossy compression

DPCM

$y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_n$

$$d_1 = y_1 \qquad\qquad d_1 \to y_1$$
$$d_2 = y_2 - y_1 \quad \xrightarrow{d_i} \quad d_2 \to y_1 + d_2 \to y_2$$
$$d_3 = y_3 - y_2 \qquad\qquad d_3 \to y_2 + d_3 \to y_3$$
$$\vdots \qquad\qquad \vdots$$
$$d_n = y_n - y_{n-1}$$

$$y_1 \to d_1 \to \hat{d}_1 \qquad\qquad \hat{d}_1 \to \hat{y}_1$$
$$y_2 \to d_2 \to \hat{d}_2 \quad \xrightarrow{\hat{d}_i} \quad \hat{d}_2 \to \hat{y}_1 + \hat{d}_2 \to \hat{y}_2$$

$\cancel{d_1}$

open loop

$d_1 = y_1 \qquad \hat{d}_1 \xrightarrow{\hat{d}_1} \hat{d}_1 \qquad \hat{y}_1 = \hat{d}_1$

$= y_1 + e_1$

$d_2 = y_2 - y_1 \qquad \hat{d}_2 \xrightarrow{\hat{d}_2} \hat{d}_2 \qquad \hat{y}_2 = \hat{d}_2 + \hat{y}_1$

$= d_2 + e_2 + y_1 + e_1$

$= y_2 + e_1 + e_2$

closed loop

$d_1 = y_1 \qquad \hat{d}_1 \xrightarrow{\hat{d}_1} \hat{d}_1 \qquad \hat{y}_1 = \hat{d}_1$

$= y_1 + e_1$

$\hat{d}_1 = \hat{y}_1$

$d_2 = y_2 - \hat{y}_1 \qquad \hat{d}_2 \xrightarrow{\hat{d}_2} \hat{d}_2 \qquad \hat{y}_2 = \hat{d}_2 + \hat{y}_1$

$\hat{y}_2 = \hat{y}_1 + \hat{d}_2$

$= d_2 + e_2 + \hat{y}_1$

$= y_2 + e_2$

subband coding