Tutorial I

19 September 2019 10:07

Lectures 1-3 Coding theory

4-13 Cryptography

CODING THEORY BASICS

→ Noisless coding theory (Huffman coding)

→ Noisy coding theory (error correcting codes)

Noisless coding theory

$$X = \{x_0, \dots, x_{n-1}\}$$

$$p_0, \dots, p_{n-1} \qquad \sum_i p_i = 1, \ p_i \geq 0$$

The GOAL: Write down outcomes of X as efficiently as possible.

(with a given alphabet Σ (here mostly $\{0,1\}$))

$x_0, \dots, x_{n-1}$

$\log_2(n)$ bit strings

$0 \to 00$
$1 \to 01$
$2 \to 10$
$3 \to 11$

is this optimal?

According to what merit? ↑ code, C

$$AVG(C) = \sum_i |c_i| \cdot p_i$$

↳ lenght of codeword corresponding to x_i

Problem formulation:

1. Given a probability distribution (random variable) design a code C with smallest average length. 2. What is the smallest achievable average length?

2.)

Shannons theorem

For a random variable $X \{p_0, \dots, p_{n-1}\}$

$$\boxed{S(X) = -\sum_i p_i \log_2 p_i \qquad \text{(Shannon entropy)}}$$

I. Average of code C for r.v. $X \geq S(X)$

II. Encoding multiple outputs together helps

III. As the number of symbols encoded together approaches infinity, the achievable average approaches $S(X)$

1a) Huffman coding

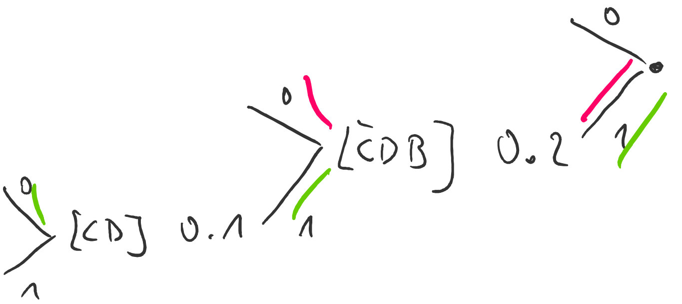
Algorithm:

INPUT: Probability distribution

OUTPUT: Code (optimal)

EX. 1.2

A	0.8
B	0.1
C	0.05
D	0.05



A → 0

B → 10

C → 110

D → 111

code

$$S(X) = -0.8 \cdot \log_2 0.8 - 0.1 \cdot \log_2 0.1 - 0.05 \cdot \log_2 0.05 - 0.05 \cdot \log_2 0.05$$

$$< 1.3$$

$$AVG(C) = [0.8] \cdot 1 + (0.1) \cdot 2 + (0.05) \cdot 3 + (0.05) \cdot 3 = 1.3$$

AA $(0.8)^2$

AB $(0.8) \cdot (0.1)$

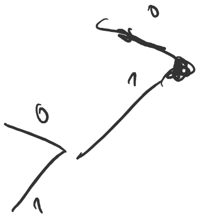
AC $(0.8) \cdot (0.05)$

⋮

A : $1/3$

B : $1/3$

C : $1/3$



A → 0

B → 10

C → 11

$$S(X) = \left[-1/3 \cdot \log 1/3\right] \cdot 3 = -\log 1/3$$

$$\approx 1.5\ldots$$

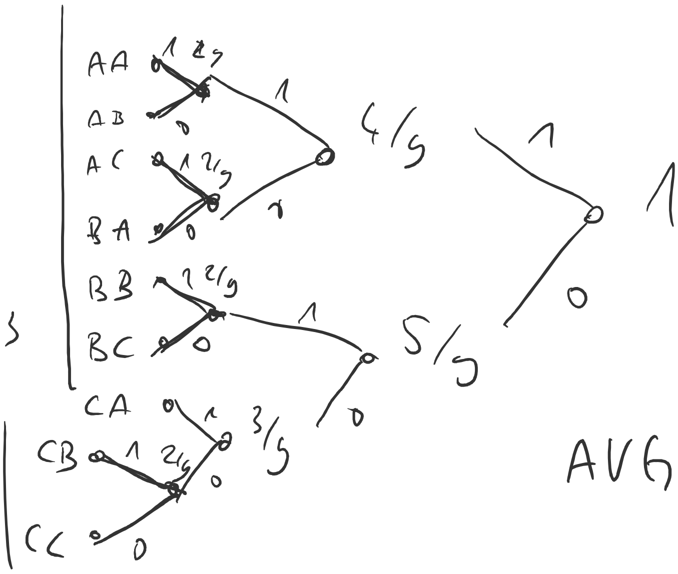
$$AVG(C) = 1/3 \cdot 1 + 2 \cdot (1/3 \cdot 2)$$

$= 5/3 = 1.666....$

AA: 1/9

Codeword of length 3

codeword of length 4



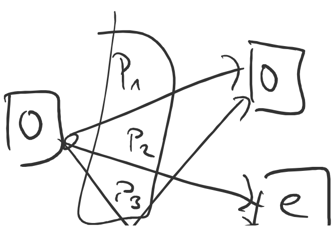
$$AVG(C) = 2\cdot\left(\tfrac{1}{9}\cdot 4\right) + 7\cdot\left(\tfrac{1}{9}\cdot 3\right) = \frac{21+8}{9} = \frac{29}{9} \approx 3.2222$$

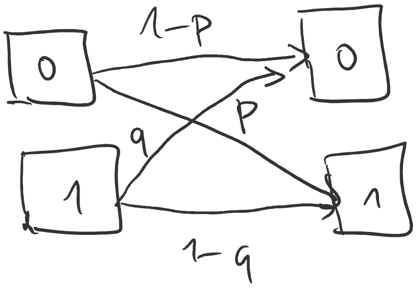
To compare we need Average per Symbol

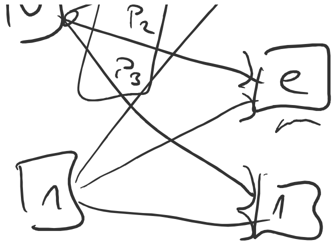
$$\frac{3.222}{2} = 1.6111....$$

As the number of repetitions n encoded together approaches ∞, then $\frac{AVG(C)}{n}$ reaches $S(X)$.

NOISY SHANNON THEORY (ERROR CORRECTING CODES)



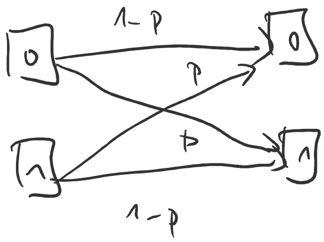




Σ	Ω
0	0
$\vdots$	$\vdots$
ℓ	k

$p(k \mid \ell)$ → probability of output k if ℓ was on the input.

Binary Symmetric channel



We have $p < 1/2$.

PRINCIPLE OF MAXIMUM LIKELY HOOD

If you observe output "0" how to interpret it?

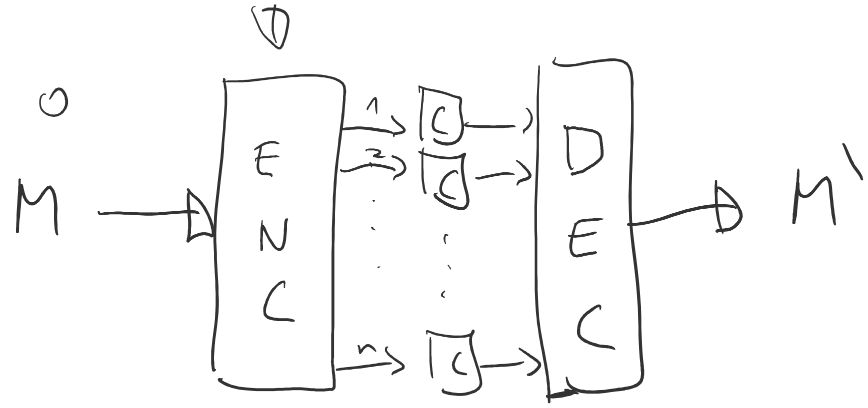
$0 \longrightarrow 0$ w.p. $1-p$

$1 \longrightarrow 0$ w.p. p

$p < 1/2 \Rightarrow 1-p > p$

Maximum likelyhood says that if you observe 0, then you should interpret it (decode it) as 0 being sent, because it is more probable of the two possibilities.

CODES



$0 \to 000$ How to decode? Use the maximum likelyhood!

$1 \to 111$

$\#1 \geq 2$ decode as 1

$\#1 < 2$ decode as 0

$000 \to 000, 001, 100, 010 \Rightarrow 0$

$111, 110, 011, 101 \Rightarrow 1$

$\boxed{01}$

$$P_r(000 \mid 001) = (1-p)(1-p)\,p$$

input output

$$\vee \quad \Leftarrow p < 1/2$$

$$P_r(111 \mid 001) = p \cdot p \cdot (1-p)$$

P_r(correct decoding to 000)

$$= (1-p)^3 + 3(1-p)^2 \cdot p > (1-p)$$

$$0 \to \underbrace{000\ldots0}_{2k+1=n}$$

$$1 \to \underbrace{111\ldots1}$$

Decoding

if $\#1 > k$ then decode 1

if $\#1 \leq k$ then decode 0

Probability of correct decoding?

Probability of correct decoding?

$$\sum_{i=0}^{k} \binom{2k+1}{i} p^i (1-p)^{2k+1-i}$$

$$\lim_{k \to \infty} \quad = 1$$

$$\boxed{\frac{\# M}{\text{length of code words}}} = \frac{2}{2k+1} \underset{k\to\infty}{=} 0$$

Code rate

Hamming distance

c_i ~> codewords $\qquad C \subseteq \{0,1\}^n$

$C = \{c_i\}$ Codes

$Ham(c_i, c_j)$ is the number of positions in which c_i and c_j differ.

$\boxed{\text{EX } 1.6}$

$\{10001, 00110, 11010, 01101\}$

$Ham(10001, 00110) = 4 \qquad Ham(00110, 11010) = 3$

$Ham(10001, 11010) = 3 \qquad Ham(00110, 01101) = 3$

$Ham(10001, 01101) = 3 \qquad Ham(11010, 01101) = 4$

5 4 3

(n, M, d)

n - length of code words
M - number of codewords
→ Code rate

d - minimum distance → determine the max. likelyhood + the number of errors code can detect or correct

Decode $\boxed{11110}$ ← 10001

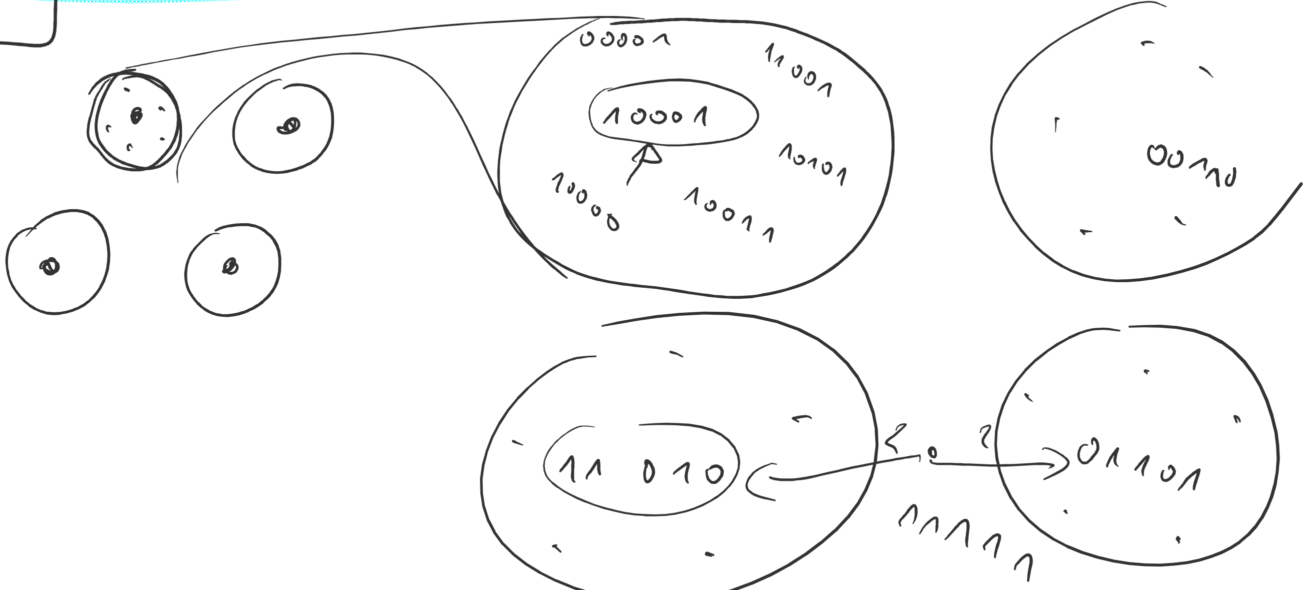
00110

11010

01101

11111

$\boxed{d = 2t+1}$ up to t errors can be corrected



n ↓
M ↑
d ↑

$\boxed{A_q(n, M)}$ largest minimum distance a code of M codewords with length n can have

over alphabet $\{0, \dots, q-1\}$

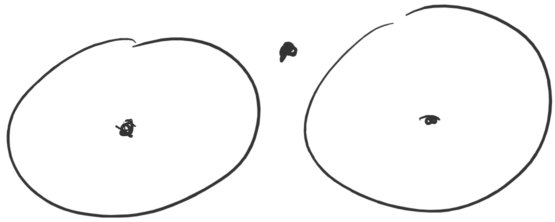
Perfect codes and sphere packing bound

$$\boxed{M\left[\binom{n}{0} + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \dots + \binom{n}{t}(q-1)^t\right] \le q^n}$$

$d = 2t+1$

M

n



$=$

Perfect codes