02 Information theory

02.01 Digital encodings

- Information
- Encoding
- Limitation
- Discretization

Information

- *Information*: reduction of uncertainty
- The minimum uncertainty is given by two alternatives
- The elementary choice between 2 alternatives contains the minimum amount of information
- *Bit*: binary digit encoding the elementary choice between 2 alternatives (*information unit*)

Words

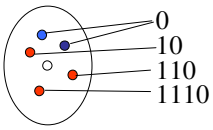
- *Word*: sequence of *N characters* (or *digits*) taken from a given finite *alphabet* of *S symbols*
- There are S^N different words of *N* characters taken from the same alphabet of *S* symbols
- There are S^N different *configurations* of *N* characters taken from an alphabet of *S* symbols
- A *binary word* is composed of *bits*, defined over the binary alphabet $B = \{0,1\}$
- *Byte*: binary word of 8 bits

Encoding

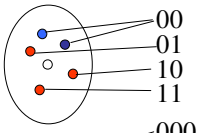
- *Encoding*: assignment of words with the elements of a set according to a given rule

Properties:

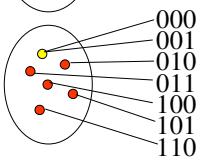
- *Irredundant*: each element is assigned with a unique word



- *Constant length*: all code words are of the same length



- *Exact*: all elements are encoded and there are no elements associated with the same word



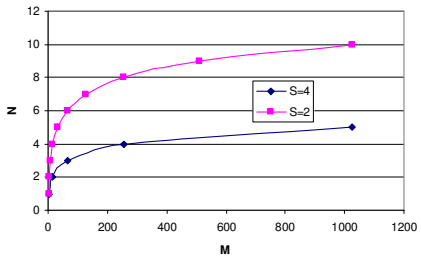
Encoding of finite sets (1)

- The minimum number of digits of a *constant-length exact encoding* of a set of *M* elements is

so that

$$N = \lceil log_S M \rceil \quad (1)$$

$$S^N \geq M \quad (2)$$



Encoding of finite sets (2)

- Properties:

$$M_1 \cdot M_2 = S^{N_1} \cdot S^{N_2} = S^{N_1+N_2} \quad (3)$$

$$log_S M_1 + log_S M_2 = N_1 + N_2 = log_S(M_1 \cdot M_2) \quad (4)$$

$$M^k = (S^N)^k = S^{k \cdot N} \quad (5)$$

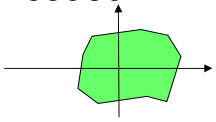
$$k \cdot log_S M = k \cdot N = log_S M^k \quad (6)$$

$$log_{S_1} M = \frac{log_{S_2} M}{log_{S_2} S_1} \quad (7)$$

N	M	
	S=2	S=4
1	2	4
2	4	16
3	8	64
4	16	256
5	32	1024
6	64	4096
7	128	16384
8	256	65536
9	512	262144
10	1024	1048576

Encoding of unlimited sets

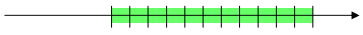
- An unlimited set contains infinite elements
  - Example: integer numbers
- Infinite sets cannot be exactly encoded

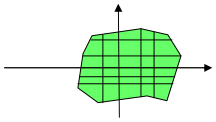


- *Limitation*:
  - In order to be digitally encoded the set must be *restricted* to a limited, finite subset
  - In most cases this is done by encoding only the elements within given lower and upper bounds
  - Example: integer numbers within 0 and 999
- The limited subset may be exactly encoded

Encoding of continuous sets

- A continuous set contains infinite elements
  - Example: real numbers in a given interval, points in a square, ...





- Discretization:
  - In order to be digitally encoded, the set needs to be *discretized*:

    partitioned into a discrete number of non-overlapping subsets covering the entire set
  - Codewords are associated with subsets
- The resulting encoding is *approximated*