CMPT 365 Multimedia Systems

Lossless Compression

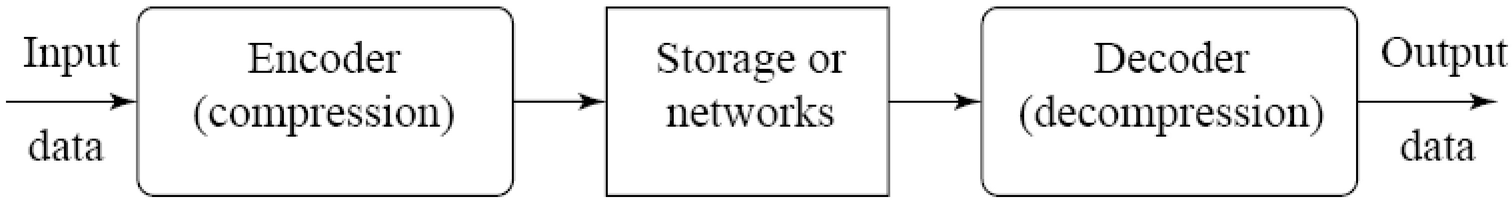
Fall 2023

Outline

- Why compression ?
- Entropy
- Variable Length Coding
  - Shannon-Fano Coding
  - Huffman Coding
  - LZW Coding
  - Arithmetic Coding

Compression

- **Compression:** the process of coding that will effectively reduce the total number of bits needed to represent certain information.



Why Compression ?

- Multimedia data are too big
  - "A picture is worth a thousand words ! "

File Sizes for a One-minute QCIF Video Clip

Frame Rate	Frame Size	Bits / pixel	Bit-rate (bps)	File Size (Bytes)
30 frames/sec	176 x 144 pixels	12	9,123,840	**68,428,800**

Approximate file sizes for **1 sec** audio

Channels	Resolution	Fs	File Size
Mono	8bit	8Khz	64Kb
Stereo	8bit	8Khz	128Kb
Mono	16bit	8Khz	128Kb
Stereo	16bit	16Khz	512Kb
Stereo	16bit	44.1Khz	1441Kb*
Stereo	24bit	44.1Khz	2116Kb

1CD 700M 70-80 mins

Lossless vs Lossy Compression

- If the compression and decompression processes induce no information loss, then the compression scheme is **lossless**; otherwise, it is **lossy**.
- **Compression ratio:**

$$compression\ ratio = \frac{B_0}{B_1}$$

B_0 – number of bits before compression
B_1 – number of bits after compression

E.g., original file of size 100KB; after compression, 20KB. Then compression ratio = 5

Why is Compression possible ?

- Information Redundancy

- Question: How is “information” measured ?

Outline

- Why compression ?
- Entropy
- Variable Length Coding
  - Shannon-Fano Coding
  - Huffman Coding
  - LZW Coding
  - Arithmetic Coding

Self-Information

Information is related to probability
Information is a measure of uncertainty (or "surprise")

- Intuition 1:
  - I've heard this story many times vs This is the first time I hear about this story
  - Information of an event is a function of its probability:

    i(A) = f (P(A)). Can we find the expression of f()?
- Intuition 2:
  - Rare events have high information content
    - Water found on Mars!!! Covid-19 case confirmed ! (Feb 2020)
  - Common events have low information content
    - It's raining in Vancouver. Covid-19 case confirmed ! (Feb 2021)

  → Information should be a decreasing function of the probability:

  Still numerous choices of f().
- Intuition 3:
  - Information of two independent events = sum of individual information:

    If P(AB)=P(A)P(B) → i(AB) = i(A) + i(B).

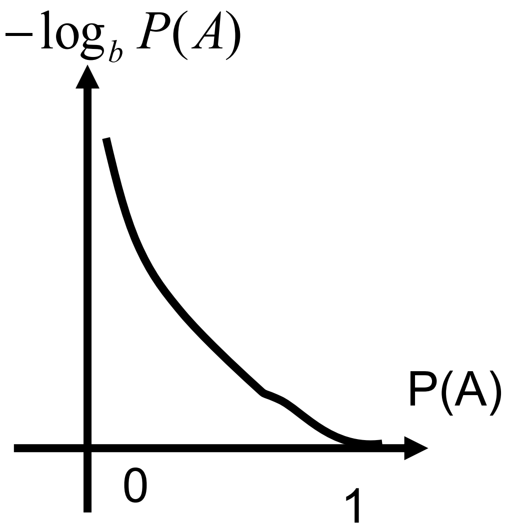
  → Only the logarithmic function satisfies these conditions.

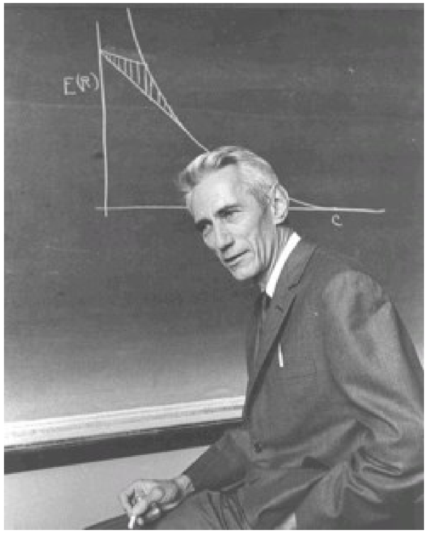
Self-information

- Shannon's Definition [1948]:
  - Self-information of an event:

$$i(A) = \log_b \frac{1}{P(A)} = -\log_b P(A)$$

If b = 2, unit of information is bits





Entropy

- Suppose:
  - a data source generates output sequence from a set $\{A_1, A_2, \ldots, A_N\}$
  - $P(A_i)$: Probability of A_i
- First-Order Entropy (or simply Entropy):
  - the average self-information of the data set

$$H = \sum_i -P(A_i)\log_2 P(A_i)$$

- The first-order entropy represents the minimal number of bits needed to losslessly represent one output of the source.

Example 1

- X is sampled from {a, b, c, d}
- Prob: {1/2, 1/4, 1/8, 1/8}
- Find entropy.

Example 1

- The entropy η represents the *average* amount of information contained per symbol in the source S

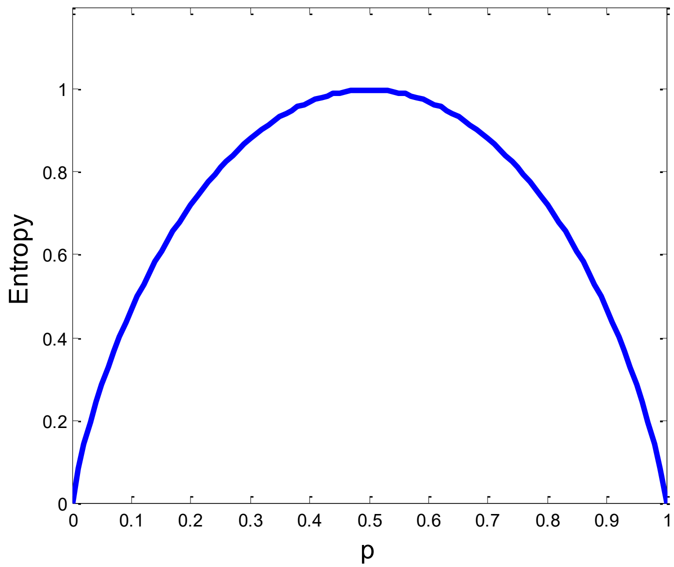
- η specifies the lower bound for the average number of bits to code each symbol in S, i.e.,

$$\eta \leq \bar{l}$$

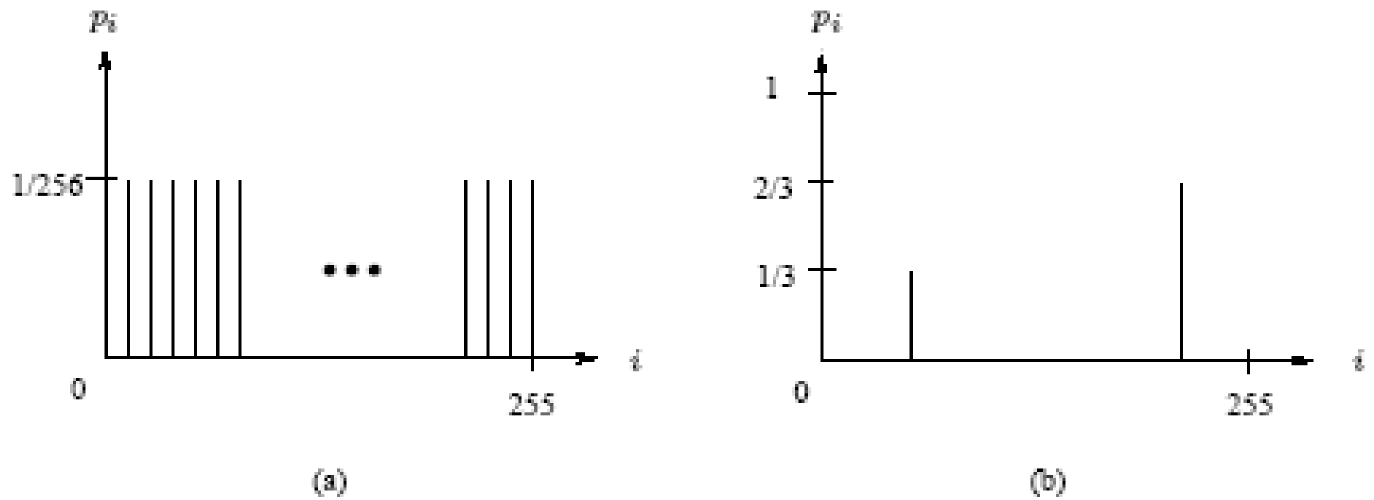
- the average length (measured in bits) of the codewords produced by the encoder.

Example 2

- A binary source: only two possible outputs: 0, 1
  - Source output example: 000101000101110101......
  - $P(X=0) = p$, $P(X=1)= 1 - p$.
- First order entropy:
  - $H = p\,(-\log_2(p)) + (1-p)\,(-\log_2(1-p))$
  - H = 0 when p = 0 or p =1
    - Fixed output, no information
  - H is largest when p = 1/2
    - Largest uncertainty
    - H = 1 bit in this case



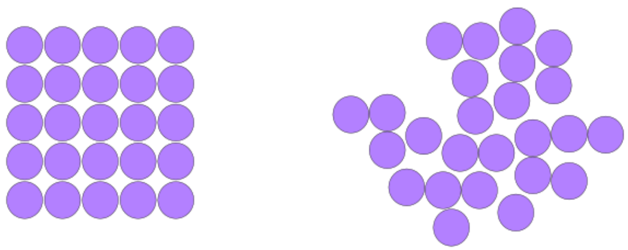
Example 3



(a) histogram of an image with *uniform* distribution of gray-level intensities, i.e., p_i = 1/256. Entropy = $\log_2 256$=8

(b) histogram of an image with two possible values. Entropy=0.92.

Entropy in Physics



Low Entropy High Entropy

The Second Law of Thermodynamics

Entropy is a measure of the disorder in a system. All systems gain entropy over time.

The Second Law of Thermodynamics says that the total entropy of both a system and its surrounding will NEVER decrease.