Describe the mathematical models used in lossless compression. Explain how probability models and physical models contribute to efficient data representation with examples.

→ Lossless compression reduces file size without losing any information.

→ It relies on mathematical models to represent data efficiently by removing redundancy.

Types

1) Probability (statistical model) models.

2) Physical (source-based) Models.

These models are used along with coding techniques such as:

* Huffman coding
* Arithmetic coding

Probability Models

They assign probabilities to symbols based on their frequency of occurence.

→ If a symbol appears frequently → assign shorter code.

→ If a symbol appears rarely → assign longer code.

This is based on shannon's information theory:

$$H(x) = -\sum p(x) \log_2 p(x)$$

where,

$H(x)$ = Entropy

$P(x)$ = probability of symbol.

Types of probability models

a) Static Model

- Probabilities fixed before compression.
- Based on known statistics.

Eg: AAAAABBC

$A = 5$, $B = 2$, $C = 1$

using Huffman coding,

A → Shortest code.

C → longest code

b) Adaptive Model.

- Probabilities updated during compression.
- suitable for unknown datastreams.

Eg: Arithmetic coding. dynamically adjusts probabilities while encoding.

Contribution to efficient representation.

→ Removes statistical redundancy.

→ Reduces average bits per symbol.

→ Achieves near entropy compression.

Eg: If probability of $A = 0.5$, minimum bits required ≈ 1 bit

instead of 8 bits (ASCII)

Thus compression is achieved.

Physical Models

→ They use knowledge of the data source structure rather than just symbol frequency.

→ They exploit patterns, repetition and context in data.

Types of physical Models

(a) Run Length Model

used when data has repeated symbols.

Eg: AAAAA

Instead of storing 5 A's → Store (A,5)

→ used in

* fax compression
* Bitmap images.

(b) Dictionary Based Model

→ Stores repeated patterns in a dictionary.

Eg: COMPUTER COMPUTER

1 → COMPUTER (instead of storing twice)

compressed output

1 1

(c) (Context Based Model)

→ Predicts next symbol based on previous symbol.

Eg: After GV → next letter likely "E".

→ used in,

- PPM → (Prediction by Partial Matching).

Contribution to efficient Representation

→ Removes structural redundancy.

→ Exploits repetition & patterns.

→ Effective for text, images and executable files.

Eg: Large blank areas in images → highly compressible using run-length encoding.